Three Essays on Forecasting and Information Acquisition in Finance

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Banking & Finance

presented by

Maria Putintseva
from Russia

approved in January 2013 at the request of
Prof. Dr. Marc Paolella
Prof. Dr. Markus Leippold

Article1: Predictive Power of Information Market Prices

Article2: Portfolio Delegation and Fiduciary Management

Article3: Mixture Normal Conditional Correlation Models

Predictive Power of Information Market Prices

This paper was presented at

- Research Seminar of the Department of Banking & Finance, Zurich, Switzerland (November 2010)
- 3d International Conference on Prediction and Information Markets, Nottingham, UK (April 2011)

Abstract
Prediction (or information) markets are markets where prices of traded contracts correspond to the probabilities of future events. Their surprisingly good predictive accuracy and manipulation resistance are pointed out by many researchers. This study verifies whether the probabilities provided by financial information markets are reliable and whether they could be fully replicated from the dynamics of an underlying asset. Additionally, the inter- and intraday microstructure of a financial prediction market of Intrade.com trading platform is examined and quantitatively described. Finally, since some ability to forecast future changes in the underlying asset is detected, a simple trading strategy based on the trading process on the considered prediction market is suggested and its profitability and applicability are evaluated.

Keywords: Prediction Market; Forecasting; Probability; Trading Strategy; Market Microstructure

JEL classification: D49; D83; G13; G14; G17

1 Introduction

The prediction (or information) market is a tool that estimates the probability of an unknown future event by aggregation of individual beliefs of traders. The market amalgamates features of financial markets and sport betting, since the delivered probabilities could be either interpreted as prices of contingent claims or mapped into odds. There are a number of aggregating mechanisms that could be applied to design such a market, but the most popular one is a double auction. In this case sellers and buyers place limit orders expressing their beliefs about an outcome, and the price at which the current bid and ask are crossing is an aggregated market belief. It is argued that this mechanism is suitable for belief aggregation, because it satisfies agent participation and preferences revelation constraints, see e.g. Pennock and Wellman, 1997.

First research works about information markets appeared in the early 90s and were mostly initiated by the successful performance of Iowa Electronic Markets, pioneers in the application of prediction market mechanism. During the following 10 years the number of new articles per year did not exceed five. The evolution of computer and communication technologies made it possible to launch online trading platforms, to store and process large amounts of data and to increase frequency of trading, which broadened the field for experiments and spurred the interest for the prediction market phenomenon. Starting from 2000, the number of research papers grows almost linearly.[1]

Online trading platforms, such as Intrade.com or Betfair.com, offer a variety of prediction markets for betting and trading: sports, politics, finance, entertainment, weather, and even exotic ones like "Whether the Higgs Boson Particle will be observed". For studying purposes it is convenient to consider contracts with an observable and measurable underlying process. The contracts traded in financial prediction markets are a good choice because they are easily interpretable as digital options, and the underlying asset data are normally available. Since studies concerned with betting on political and sport events are prevailing in the field, this research work is aimed to offset and pursues a quantitative study of financial prediction markets. The inspiration comes partly from the findings of Tetlock, 2008, who analyses the data obtained from TradeSports and concludes that financial prediction markets are surprisingly efficient as opposed to sport betting markets. He states that a prediction market with higher liquidity does not necessarily provide better calibrated prices.

The current paper examines the following aspects of the Intrade.com prediction market for

[1] For more details see Figure 2 in Tziralis and Tatsiopoulos, 2007, an extended literature review of about 15 years of prediction market research history published in the first issue of The Journal of Prediction Markets.

the Dow Jones Industrial Average (DJIA) index. First, a quantitative analysis of market microstructure and market liquidity is conducted. Several important regularities of the trading process have been recognized. Second, some empirical evidence is provided for the fact that the prices from the prediction market better reveal the true probability. It seems impossible to fully replicate the information incorporated in them using only the values of the DJIA index. Third, a polynomial approximation for the price function of the digital option is developed and estimated. This approximation also allows for the derivation of the state price density. Finally, an application of the information from the considered prediction market to trade the DJIA index is suggested and tested.

The research done in Zitzewitz, 2006 is to a great extent in line with my study. The author focuses his attention on the development of a practical application of prediction market information. He scrupulously analyses the main features of the TradeSports DJIA daily and intraday digital option markets, constructs an intraday implied volatility measure using option prices and confirms its incremental predictive power for changes in future realized volatility and the DJIA futures. To my knowledge, there is no publicly available research investigating the microstructure of Intrade.com prediction market, which is now claimed to be the world leader in the branch after TradeSports.com ceased its operations in November 2008. The only research work providing a practical application of the data from Intrade.com is that of Archak and Ipeirotis, 2008. They develop a model for the volatility of prediction market prices, which performs well compared to the GARCH model and could be used for option pricing. Another example of the empirical study of the financial prediction markets is Gürkaynak and Wolfers, 2006. They consider the market for "Economic Derivatives" launched by Goldman Sachs and Deutsche Bank and tied to the macroeconomic outcomes. The auction mechanism is pari-mutuel betting with limit orders, which is unusual for financial markets, but it helps to resolve the liquidity problem. The authors derive state pricing functions for different economic variables and find that the forecasts of this market are more precise than those of experts and risk premia in the prices are extremely small for the most plausible utility functions.

A large part of the literature on prediction markets is investigating their properties. Wolfers and Zitzewitz, 2004 summarize characteristics, benefits and drawbacks of prediction markets detected within the first decade of their exploitation. In particular, they argue for accuracy of information aggregation, scarcity of arbitrage opportunities and manipulation resistance, but warn about the presence of a "favorite-longshot" bias, subjective trading and speculative bubbles. Wolfers and Zitzewitz, 2006 explore conditions under which the prediction market prices correspond to average beliefs and show that these conditions are fulfilled for a broad range of distributional assumptions. Berg et al., 2008 present the empirical evidence of a better long-term forecasting using prediction markets compared to polls by analysing the election market data gathered from Iowa Electronic Markets. Substantial interest is also being attracted to the issues of prediction market design, efficiency of person-to-person and parimutuel betting, and to theoretical explanations of prediction market features, e.g. Ottaviani and Sørensen, 2008 thoroughly investigate favorite-longshot bias origins.

To summarize, the results presented in this work support the thesis that prediction markets, in spite of being dominated by stock exchange giants, are able to correctly predict probabilities of future events. They also possess dynamic regularities and could be used to study trader's beliefs and to design trading strategies. The paper is organized as follows. Section 2 describes some important features of the considered market and summarizes the statistical properties of the data. Section 3 focuses on market microstructure and regularities, Section 4 is devoted to the analysis of the option price correctness. Section 5 suggests an algorithm of the pricing function estimation. Section 6 shows a practical application of prediction market information, Section 7 concludes.

2 Intrade DJIA Market and Data

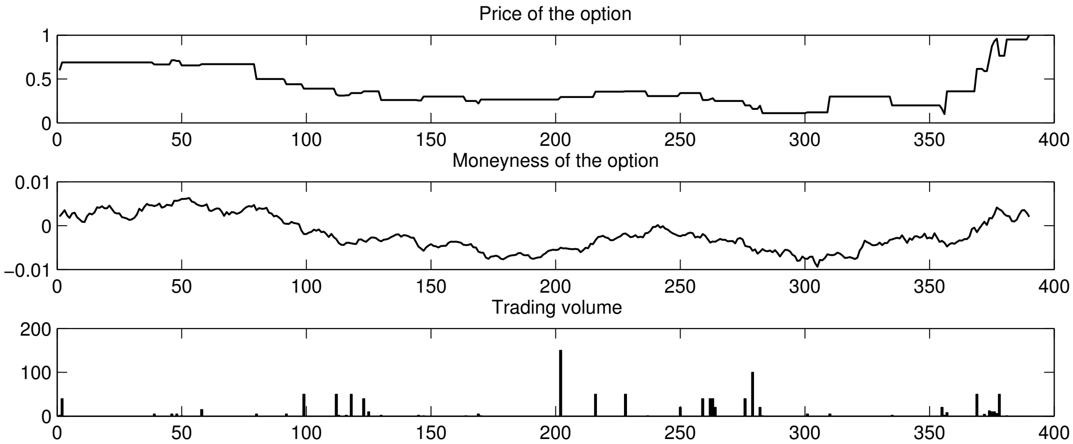
The data available at the beginning of the current research consist of minute-by-minute values of the Dow Jones Industrial Average (DJIA) index and records of trades of the digital options on the Index taken from the online trading platform Intrade.com. In general, it is not correct to consider the DJIA index itself: since it cannot be bought or sold, one should take DJIA futures or a DJIA exchange-traded fund. A lack of these data forces to assume that the DJIA index is not deviating much from a fund replicating it and that the transaction costs are negligible.

Contract specification is the following: the contract will expire at 100 points ($10 dollars) if the DJIA index closes or prints higher at the end of trading day than the previous day's close. The contract will expire at 0 points if it does not close or print higher. Therefore, the considered digital option lives only one day. Each trade record includes the price of the market order and the volume traded at that price. The maximum available data frequency for this market is one minute.

The time span of the whole data sample is from March 9, 2009 till January 29, 2010, excluding weekends, holidays and some other missing days (e.g., on the last week of December

2009 the trading platform was open, but there was no market maker and, therefore, almost no activity). The standard trading day is 390 minutes long (from 9:31 till 16:00), and I do not consider trades beyond this time interval. Figure 1 shows a common picture observed on the market during the day. All in all, there are 85410 observations for the Index and 10958 trade records for the digital options.

Figure 1: Intraday market dynamics on July 8th, 2009



Note: The first graph from the top presents the price of the digital option on the DJIA index normalized to one. The second graph is the moneyness of the option (the current log price of the DJIA index minus its log previous close price), the lowest graph shows the corresponding trading volume. On the X-axis are minutes.

Strictly speaking, there is no unique price on the market except at the very moment of negotiation. Therefore, one needs to apply some interpolation procedure to derive the price between trades. The first option is to connect two consequent prices with a piece of a linear function, the second option is to keep the last price until a new trade occurs. Here the latter option is used, but experiments show that the choice of a method has negligible impact on results.

There is an independent market maker, who has an agreement with Intrade.com to provide consistent liquidity of a minimum size in return for reduced participation fees, but no profit-sharing agreement. Only the market maker is allowed to use an automated program for pricing. The option to become a market maker is open to any registered member of the trading platform, provided that he or she is willing and able to supply the required level of liquidity.

During the period of data accumulation there were two types of fees for traders: the trading fee for a price-taker of \$0.03 (\$0.05 for extreme prices) and the "In the Money Expiry" fee of \$0.10 per lot. As of the 1st January 2011 a new policy is applied: all members are charged the "trading seat" fee of \$4.99 per month, and all other fees are cancelled.

Table 1 displays some statistics describing market variability during the year. The first column containing monthly averaged daily trading volume provides the evidence of seasonal patterns in trading activity. Indeed, there are two spikes of traded volume in March and September, the months when half-year company reports are made public, and a substantial drop in activity during July, August, December and January (usual time for summer and winter holidays). Mean daily prices and their standard deviations, averaged within a month (the second and third columns) are very uniform across all months. The last column of monthly shares of contracts expired at 100 points displays some bias to positive trading outcomes and it could be tested whether the hypothesis "the share of positive outcomes is 0.5" is rejected or not. The z-score from the data is 2.1655. Applying the bootstrap procedure, one can calculate critical values of this statistics: $Q_{0.95} = 1.6609$ and $Q_{0.99} = 2.3846$. Therefore, the hypothesis is rejected at the 5% significance level, but not rejected at the 1% significance level.

Table 1: Summary statistics of trading seasonality

	Average Daily Volume	Daily Mean of Price, averaged	Daily S.D. of Price, averaged	Share of Contracts Expired at 100
Mar 09	1,095.8	0.5345	0.1547	0.5882
Apr 09	644.2	0.5254	0.1423	0.5714
May 09	702.1	0.5240	0.1473	0.5000
Jun 09	751.7	0.5166	0.1475	0.4545
Jul 09	491.1	0.5374	0.1327	0.7273
Aug 09	380.9	0.4971	0.1286	0.6190
Sep 09	1,044.6	0.5437	0.1600	0.5238
Oct 09	650.3	0.5296	0.1303	0.5000
Nov 09	450.9	0.5777	0.1170	0.6842
Dec 09	293.2	0.5488	0.1321	0.6111
Jan 10	340.5	0.4717	0.1567	0.5263

Note: The table summarizes seasonal differences in trading activity. The first column contains the daily trading volume, averaged within a month. The second and third column show the mean and the standard deviation of the option price, averaged within a month. The last column contains monthly shares of positive outcomes of the underlying event.

Table 2 illustrates the distribution of trading activity across different values of moneyness and time. The measure of activity is a ratio of the number of trading minutes to the total number of minutes in an interval given that moneyness lies in a certain range. In the first trading hour the activity is almost uniformly distributed across the values of moneyness meaning that traders who take their initial positions are not averse to trade even if the current situation on the market does not look promising. As time passes by, the activity concentrates on the values of moneyness, which are close to zero. It is mostly noticeable during the last one and a half trading hours. Clearly, if one bought the option at the beginning of the day and the moneyness is still high, one can simply wait until the option expiry being inactive.

Table 2: Trading activity for different values of time and moneyness

	Moneyness at the moment of trading					
Time interval	0.01 or more	0.005 to 0.01	0 to 0.005	-0.005 to 0	-0.01 to -0.005	-0.01 or less
9:31-10:30	0.1506	0.1949	0.1867	0.1802	0.1971	0.1356
10:31-11:30	0.0619	0.1049	0.1070	0.1255	0.1267	0.0702
11:31-12:30	0.0257	0.0888	0.0985	0.1070	0.0975	0.0327
12:31-13:30	0.0247	0.0742	0.1225	0.1278	0.0830	0.0218
13:31-14:30	0.0315	0.0722	0.1612	0.1443	0.0805	0.0226
14:31-15:30	0.0267	0.0952	0.3449	0.2519	0.0940	0.0363
15:31-16:00	0.0571	0.1115	0.5397	0.5337	0.1136	0.0949

Note: The table shows the distribution of trading activity in the "time-moneyness" space. The activity is measured as a ratio of trading minutes to the total number of minutes corresponding to particular values of time and moneyness.

The last important thing to look at is the distribution of volume across different values of moneyness and time presented in Table 3. The volume is averaged within particular values of time and moneyness, presented in the table. Again, volume could be considered to be uniformly distributed at the beginning of the day. In the second trading hour, in the middle of the day and in the last 30 minutes there is a shift towards highly negative values of moneyness that might indicate the periods in which a trader analyses the current situation and if she is not sure about the positive outcome, she liquidates her position. About one hour before closing there is a shift towards highly positive values of moneyness, meaning that some traders just implement low-risk strategies of buying the option shortly before the expiration despite its probably high price hoping to get some small gains.

Table 3: Average volumes for different time and moneyness

	Moneyness at the moment of trading					
Time interval	0.01 or more	0.005 to 0.01	0 to 0.005	-0.005 to 0	-0.01 to -0.005	-0.01 or less
9:31-10:30	12.94	10.15	13.59	14.96	14.55	11.51
10:31-11:30	14.78	12.63	12.44	11.52	11.42	18.64
11:31-12:30	10.51	10.51	14.07	12.43	15.07	14.56
12:31-13:30	11.40	10.82	11.44	14.82	13.29	25.21
13:31-14:30	17.31	10.12	10.61	18.39	14.02	17.07
14:31-15:30	18.69	11.54	12.62	13.39	15.90	12.51
15:31-16:00	13.24	9.99	8.73	12.88	9.79	24.80

Note: The table shows the distribution of trading volume in the "time-moneyness" space. The volume is averaged within particular values of time and moneyness.

3 Market Microstructure

The considered prediction market is not deep: the sample average daily trading volume is 615, the minimum trading volume is 2 and the maximum trading volume is 3124. The market is also not very active: the trading minutes make up only about ten percent of the whole sample of minutes. In this regard it could be useful to find answers to the following questions:

1. What factors influence the probability of a trade to occur?

2. What characteristics of the market promote large rapid changes in the option price?

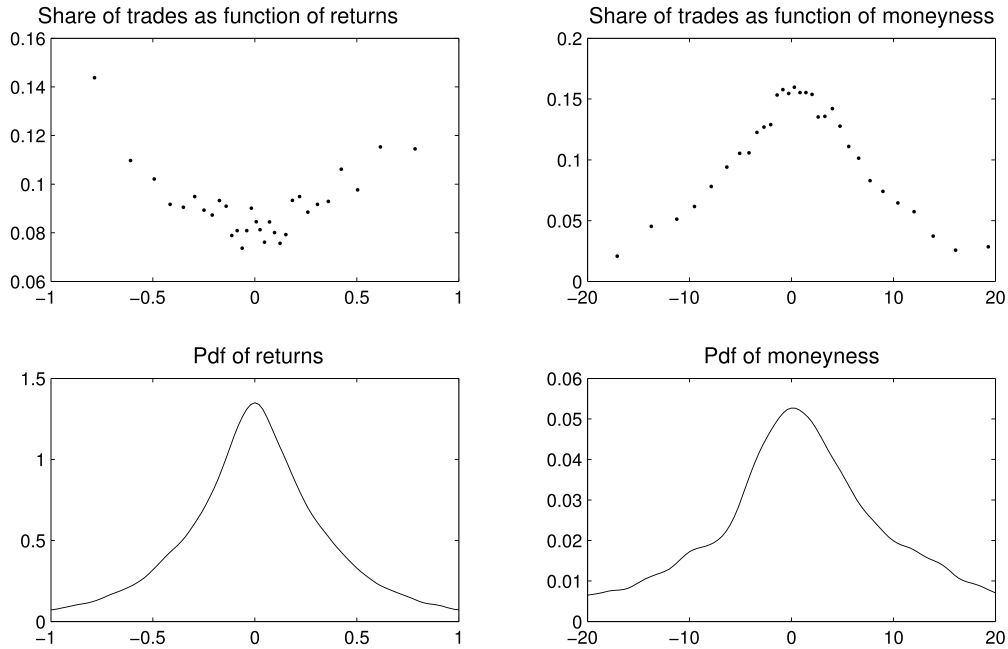
3. Is it possible to trade large volumes with little or no impact on the option price?

This information can help to understand how beliefs react to new information and to design a trading strategy for this market.

Visual examination of the joint dynamics of the option price $P_{j,t}$ and the value of DJIA index $S_{j,t}$ suggests that the following characteristics have the most prominent effect on trading activity: return $r_{j,t} = 1000(\log S_{j,t} - \log S_{j,t-1})$ and moneyness $M_{j,t} = 1000(\log S_{j,t} - \log C_{j-1})$, both multiplied by 1000 for computational convenience. The indexes j and t denote the day and the minute respectively. The influence of returns is intuitively clear, since they incorporate arriving information to which the option prices may react immediately or with a lag (or may not react at all). As for moneyness, the more the current DJIA index value exceeds the previous closing price, the greater is the probability that the option will expire at 100 points.

The upper part of Figure 2 shows how the share of trading minutes depends on returns and moneyness of the Index and the lower part shows non-parametrically estimated densities of these factors. The activity is concentrated on large returns and small absolute moneyness, and the latter has much more dramatic impact (and occurs more often). Since there is no noticeable asymmetry in the impact of positive/negative returns and moneyness, it is reasonable to consider further their absolute values.

Figure 2: Relation between trading activity, returns and moneyness



Note: The upper graphs show the distribution of trading activity across returns and moneyness. Trading activity is measured as a ratio of the number of trading minutes to the total number of minutes, which correspond to the values of return/moneyness lying in certain ranges. These ranges are defined by the values of empirical quantiles of returns/moneyness with quantile order increment of 0.03. The densities of return and moneyness on the bottom graph are non-parametrically estimated using the standard normal kernel function.

To answer the first question I use the probit regression to fit the conditional expectation of the binary indicator of a trade in the current moment t $A_{j,t}$. Complementary to such important explanatory variables as contemporaneous and lagged absolute returns $|r_{j,t}|$ and $|r_{j,t-1}|$ of the DJIA index, absolute moneyness $|M_{j,t}|$ and time of trade t, I include in the list of regressors the log volume traded in the last trade $V_{j,\tau-1}$, duration $D_{j,t}$ (the number of minutes since the last trade) and the dummy variable indicating trade in the previous period $A_{j,t-1}$.

Table 4: Probit regressions for trading activity

	Const	t	$\|r_{j,t}\|$	$\|r_{j,t-1}\|$	$\|M_{j,t}\|$	$V_{j,\tau-1}$	$A_{j,t-1}$	$D_{j,t}$
(1)	-1.5442	0.0002	0.3097	0.2176	-0.0037			
t-stat	(-105.61)	(4.43)	(21.25)	(14.74)	(-6.92)			
(2)	-1.3164	0.0007	0.2721	0.1830	-0.0306	0.0167	0.3872	-0.0102
t-stat	(-71.35)	(12.69)	(17.58)	(11.72)	(-24.06)	(3.66)	(21.78)	(-25.09)

Note: The specification of the probit regression is the following: $\mathbb{E}(A_{j,t} = 1|X) = \Phi(\beta X)$, $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$. The dependent variable is the binary indicator of trade in the current moment t $A_{j,t}$. The regressors are: the contemporaneous return $|r_{j,t}|$ of the Index, the lagged return $|r_{j,t-1}|$ of the Index, absolute moneyness $|M_{j,t}|$, the time of trade t, the log volume traded in the last trade $V_{j,\tau-1}$, duration (the number of minutes since the last trade) $D_{j,t}$, the binary indicator of trade in the previous period $A_{j,t-1}$. All coefficients are significant at the 0.1% significance level.

In Table 4 the estimation results for simple and extended specifications are presented. All coefficients are significant at the 0.1% level. It follows from the simple regression that, first, the larger is the amplitude of the DJIA value fluctuations, the higher is the trading probability. Second, consecutive large price changes increase the trading probability irrespective of their signs. Third, if the price of the DJIA index goes too far away from the previous close price regardless of direction, the trading activity calms down. Fourth, the activity increases as the end of the trading day approaches.

The extended specification adds the last three variables from the list to the pool of regressors. Consideration of the results leads to several conclusions. First, the higher is the volume of the last trade, the higher is the probability of trade occurrence. There are at least two reasons for that. (1) Low volume may mean that the trade was rather accidental than informative and could be ignored. (2) High volume may indicate the beginning of a "trading wave": since the market is not deep, it is problematic to buy or sell a desired volume if it is relatively large, therefore, traders have to split their orders. Second, the fact that there was also a trade in the previous period increases the probability of a trade in the current period. It indicates that trades arrive in clusters. It could be explained by (1) order splitting or (2) differentiation between important and unimportant information. It means that there is no trade if arriving information cannot influence the current estimate of the outcome probability or it has been already forecast and incorporated in the price; but if important information arrives, there is an adjustment taking some time. Finally, the impact of duration is negative, which means that the longer the market does not move, the higher is the probability that it will not move in the future. It could be considered from a psychological aspect: each trader is afraid to make a market order, because others will use his signal and have information gains. It is also related to the argumentation about trade clustering.

Now let us look for the main determinants of both large absolute changes in the option price and trades occurring without any price change. The price is normalized to one and the threshold of 0.1 is taken to cut off large changes, which corresponds to 8.5% of all trades. Trades without price changes account for 14.5% of all trades. Two probit regressions are fitted, with dependent variables being binary indicators of a large change or no change in the price. As before, two different specifications are considered, a simple one, which includes a constant, the trading time, contemporaneous and lagged absolute returns and absolute moneyness, and an extended one, where the log volume of the previous trade, duration and the absolute price change in the previous trade are added.

Table 5: Probit regressions for large price changes and price stability

	Const	t	$\vert r_{j,\tau}\vert$	$\vert r_{j,\tau-1}\vert$	$\vert M_{j,\tau}\vert$	$V_{j,\tau-1}$	$D_{j,\tau}$	$\vert\Delta P_{j,\tau-1}\vert$
dependent variable $\mathbf{1}_{\vert\Delta P_{j,\tau}\vert>0.1}$								
(1)	-1.3322	0.0006	0.3426	0.1248	-0.0969			
t-stat	***(-28.50)	***(4.76)	***(11.30)	***(3.79)	***(-15.13)			
(2)	-1.3972	0.0006	0.3746	0.1579	-0.1250	0.0031	0.0109	0.5825
t-stat	***(-26.24)	***(4.20)	***(12.01)	***(4.66)	***(-17.39)	(0.24)	***(13.62)	**(3.06)
dependent variable $\mathbf{1}_{\Delta P_{j,\tau}=0}$								
(1)	-1.1869	0.0000	0.0175	0.0094	0.0219			
t-stat	***(-31.50)	(0.30)	(0.62)	(0.31)	***(7.67)			
(2)	-0.9665	0.0000	-0.0278	-0.0358	0.0383	-0.0478	-0.0292	-0.7018
t-stat	***(-21.05)	(0.26)	(-0.95)	(-1.15)	***(12.04)	***(-4.27)	***(-11.99)	***(-3.46)

Note: The specifications of the probit regressions are the following: $\mathbb{E}(\mathbf{1}_{|\Delta P_{j,\tau}|>0.1}|X) = \Phi(\beta_1 X)$ and , $\mathbb{E}(\mathbf{1}_{\Delta P_{j,\tau}=0}|X) = \Phi(\beta_2 X)$, $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$. The dependent variables are the binary indicators of price changes $\mathbf{1}_{|\Delta P_{j,\tau}|>0.1}$ or the absence of change $\mathbf{1}_{\Delta P_{j,\tau}=0}$. The regressors are: the contemporaneous return $|r_{j,t}|$ of the Index, the lagged return $|r_{j,t-1}|$ of the Index, absolute moneyness $|M_{j,t}|$, the time of trade t, the log volume traded in the last trade $V_{j,\tau-1}$, duration (the number of minutes since the last trade) $D_{j,t}$, the absolute price change in the last trade $|\Delta P_{j,\tau-1}|$. t-statistics are given in parentheses. The levels of significance are the following: * is 5%, ** is 1% and *** is 0.1%.

Table 5 shows the results. When the simple specification is extended, all significant variables remain significant. Absolute moneyness is the factor of both large changes (a negative impact) and absence of changes (a positive impact). Absolute returns and the trading time do not have significant impact on the probability of a trade without price change, but are positive factors of large changes with the impact of the contemporaneous return being more than two times higher than that of the lagged return. In the extended specification duration and the absolute change of the price in the previous trade are significant and have a positive impact on the probability of large price changes and a negative impact on the probability of no change. The log volume of the previous trade is the significant factor only for the probability of no change and has negative sign. To draw a conclusion, first, one can expect large price changes if the option is nearly at-the-money, a previous or a current change of the price of either the option or the DJIA index is large and/or there is no activity on the prediction market for quite a long time. Second, the price is likely to remain the same if the option is in- or out-of-the money and/or small volumes are liquidated with small intervals between trades (order splitting).

The last question on the list is the impact of traded volume on the option price. The literature on market microstructure mentions (see e.g. Berry and Howe, 1994) the intraday U-shaped pattern in the market activity. Therefore, one should be aware that the explanatory variables could exhibit a changing impact on the option price during the day, therefore one always has to control for the trading time while running regressions. To deal with this phenomenon in another way, I split the trading day into seven intervals, six of one hour length and the last one of half an hour length. Then for each time interval one can run regressions using the within-interval averages of variables with observations pooled across the trading days. The average is taken in the number of trades within the time interval Δt on a particular day j. The observations with no trading activity in the time interval are removed, so the number of observations across intervals varies from 216 to 156.

The dependent variable is the average option price $\overline{P_{j,\Delta t}}$; the explanatory variables are the average absolute return $\overline{|r_{j,\Delta t}|}$, average moneyness $\overline{M_{j,\Delta t}}$ and the average "signed volume" $\overline{\Upsilon_{j,\Delta t}} = \overline{V_{j,\Delta t} \cdot \mathrm{sgn}(\Delta P_{j,\Delta t})}$. The "signed volume" allows for distinguishing between trades initiated by a buyer or a seller. Unfortunately, there is no such information in the data, therefore it is assumed that the trade that leads to an increase in the option price is initiated by a buyer and the trade that leads to a decrease in the option price is initiated by a seller. This approach is used, in particular, by Lee and Ready, 1991. If the "signed volume" is positive, it means that the buying activity is more intensive and vice-versa. In the regressions below positive and negative "signed volumes" are considered separately, which allows for drawing some conclusions about the impact of selling/buying activity at different times of the day.

Table 6: Intraday dependencies between option price and traded volume

	9:31-10:30	10:31-11:30	11:31-12:30	12:31-13:30	13:31-14:30	14:31-15:30	15:31-16:00
$Const$	0.5187	0.4677	0.4963	0.5224	0.5145	0.5477	0.6361
	***(34.22)	***(29.58)	***(30.73)	***(25.19)	***(25.93)	***(17.21)	***(19.92)
$\overline{\lvert r_{j,\Delta t}\rvert}$	-0.0288	0.0543	-0.0426	-0.1127	-0.0299	-0.0780	-0.1599
	(-1.38)	(1.94)	(-1.15)	*(-2.17)	(-0.77)	(-1.31)	**(-3.11)
$\overline{M_{j,\Delta t}}$	0.0273	0.0266	0.0272	0.0287	0.0320	0.0283	0.0244
	***(-32.60)	***(-29.49)	***(-27.55)	***(-21.11)	***(-22.17)	***(-15.56)	***(-14.07)
$(\overline{\Upsilon_{j,\Delta t}})^+$	-0.0002	0.0023	0.0021	0.0016	-0.0003	0.0003	0.0054
	(-0.15)	*(2.42)	*(2.27)	(0.81)	(-0.33)	(0.13)	(1.22)
$(\overline{\Upsilon_{j,\Delta t}})^-$	0.0000	0.0001	0.0005	0.0011	0.0023	0.0026	0.0100
	(0.01)	(0.06)	(0.54)	(1.53)	(0.98)	(1.19)	**(2.94)
$\overline{V_{\Delta t}}$	129.06	77.96	60.21	64.63	76.69	124.78	81.21
$\overline{N_{\Delta t}}$	9.65	6.09	4.64	4.86	5.57	9.49	8.29
Obs.	216	200	183	172	166	156	189

Note: $\{j, \Delta t\}$ is the joint index of an observation. The dependent variable is the average option price $\overline{P_{j,\Delta t}}$; the explanatory variables are the average absolute return $\overline{|r_{j,\Delta t}|}$, average moneyness $\overline{M_{j,\Delta t}}$ and the average "signed volume" $\overline{\Upsilon_{j,\Delta t}} = \overline{V_{j,\Delta t} \cdot \mathrm{sgn}(\Delta P_{j,\Delta t})}$. The averages are taken across the trades in time interval Δt of the day j and pooled across the days. t-statistics are given in parentheses. Levels of significance are the following: * is 5%, ** is 1% and *** is 0.1%.

The results are presented in Table 6. The second and the third rows from the end give the evidence of the U-shaped curves corresponding to the average trading volume and the average number of trades (the numbers in the last column should be multiplied by 2, because the length of the interval is only 30 minutes). Basically, in each time period the option price is determined by a constant level which is close to 0.5 and moneyness, which means that in the absence of any information the market in general does not give any preference to the positive or negative final outcome.

The average absolute return could be interpreted as a measure of realized volatility. It could be seen that the coefficient of this regressor is negative and significant at the 5% level in the time interval from 12:31 till 13:30, and negative and significant at the 1% level in the last half an hour of the day. It means that the volatility in these periods is perceived as bad news and its rise leads to an option price decrease, especially during the last trading minutes.

The impact of volume on the price can be barely detected from the sample. The results only demonstrate that if one tries to buy large volumes during the time period from 10:30 to 12:30 so that the average "signed volume" becomes or stays positive, then the option price will rise. On the contrary, massive selling in the last trading 30 minutes makes the price fall. It means that it might be difficult to take a large position at the beginning of the day for a desired price if the market believes in the positive outcome and to liquidate a large position at the very end of the day if the situation on the DJIA market fluctuates dramatically.

4 Quality of Market Price

The prices of digital options traded on information markets should reveal the probability of the future outcome of the corresponding event. Probability is a quite ephemeral notion, which is not observable, therefore, it is not so easy to answer the question whether prices on information markets reflect the true probability of an outcome, because nobody knows what the true probability exactly is. In asset pricing one generally believes that the prices of securities and derivatives provided by the market are correct, and relying on them the prices of other products are calculated. Here the informational quality of the price itself is doubted.

One possible solution would be to derive these probabilities in an alternative way and to compare them to the original ones. Since one also observes the value of the DJIA index, one can derive the probability of the positive outcome at any moment of time from its dynamics making some assumptions. The crucial assumption is that one can neglect changes in the distribution of the Index during a short period of time, therefore, one can simulate other possible paths of its price process by drawing randomly with replacement from the realized path.

To do so, the block bootstrap procedure is applied, because it allows for preserving the time structure of the sample (for details see e.g. Lahiri, 2003). The length of the block is set to be 15 observations (a quarter of an hour). Draws are made from 780 (2 trading days) observations preceding the current moment of time excluding draws that consist of observations from two different days. The procedure of getting the alternative value of outcome probability is the following.

1. Fix the moment of time at which the probability (the price of the digital option) should be calculated.

2. Drawing blocks from the previous 780 observations, simulate a path of the DJIA index up to the moment of the contract expiration. Record the price of the option at the expiry date.

3. Repeat the path simulation 1000 times (increasing this number further does not change the results) and average the recorded prices. This value is the probability calculated model-free from the DJIA index.

Now it would be interesting to compare these two samples of prices: the one provided by the market, and the other derived from the information extracted from the Index. Potentially, the former should contain the same information as the latter, but does it also contain any additional information? To answer this question, one has to compare the abilities of these two prices to correctly predict the outcome and to find out, which of them is closer to the true probability.

The probit regression of $P_{j,T}$ (the price of the option at the expiration) on the probability of the normalized option payoff to be equal to 1 estimated at the moment of time t could be written as

$$\mathbb{E}(P_{j,T} = 1|P_{j,t}) = \Phi(\beta_{0,t} + \beta_{1,t}P_{j,t}), \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz. \tag{1}$$

Ideally, a change in the information incorporated in the conditional expectation of $P_{j,T}$ should be one-to-one incorporated in $P_{j,t}$:

$$\frac{\partial \mathbb{E}(P_{j,T} = 1|P_{j,t})}{\partial P_{j,t}}(\bar{P}_{j,t}) = \phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t} = 1, \quad \phi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{2}$$

If the marginal effect $\phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t} > 1$, then $P_{j,t}$ underreacts, and if $\phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t} < 1$, then $P_{j,t}$ overreacts. The probit model in (1) is usually estimated by the maximum likelihood method, which makes it easy to derive the asymptotic covariance matrix $\Sigma_{\beta,t}$ of parameters $(\beta_{0,t}, \beta_{1,t})$. Then we can derive the asymptotic variance of $\phi(\beta_{0,t}+\beta_{1,t}\bar{P}_{j,t})\beta_{1,t}$ by means of the delta method (see e.g. Greene, 2003, p.70):

If $\sqrt{n}(z_n - z) \xrightarrow{d} N(0, \Sigma)$, *where* z_n *is a sequence of* k*-dimensional random vectors,* g *is a function such that* $g : \mathbb{R}^k \to R^l$ *and continuously differentiable at* z, *then* $\sqrt{n}(g(z_n) - g(z)) \xrightarrow{d} N(0, G\Sigma G')$, $\quad G = \frac{\partial g(u)}{\partial u'}|_{u=z}$.

Differentiating $\phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t}$ with respect to $\beta_{0,t}$ and $\beta_{1,t}$, one gets

$$\frac{\partial \phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t}}{\partial \beta_{0,t}} = -\phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t} = G_t(1), \tag{3}$$

$$\frac{\partial \phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t}}{\partial \beta_{1,t}} = \phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})(1 - (\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t}\bar{P}_{j,t}) = G_t(2). \tag{4}$$

Therefore, the marginal effect has the following asymptotics

$$\sqrt{J}\left(\phi(\hat{\beta}_{0,t} + \hat{\beta}_{1,t}\bar{P}_{j,t})\hat{\beta}_{1,t} - \phi(\beta_{0,t} + \beta_{1,t}\bar{P}_{j,t})\beta_{1,t}\right) \xrightarrow{d} N(0, G_t\Sigma_{\beta,t}G_t'), \quad G_t = [G_t(1), G_t(2)], \tag{5}$$

and one could use the statistic

$$\tau = \sqrt{J} \frac{\phi(\hat{\beta}_{0,t} + \hat{\beta}_{1,t}\bar{P}_{j,t})\hat{\beta}_{1,t} - 1}{\sqrt{\hat{G}_t \Sigma_{\hat{\beta},t} \hat{G}_t'}} \tag{6}$$

to test the hypothesis whether the marginal effect is equal to 1 for different values of $\bar{P}_{j,t}$. Under the null hypothesis the statistic τ is asymptotically distributed as the standard normal distribution. J is the number of days in the sample.

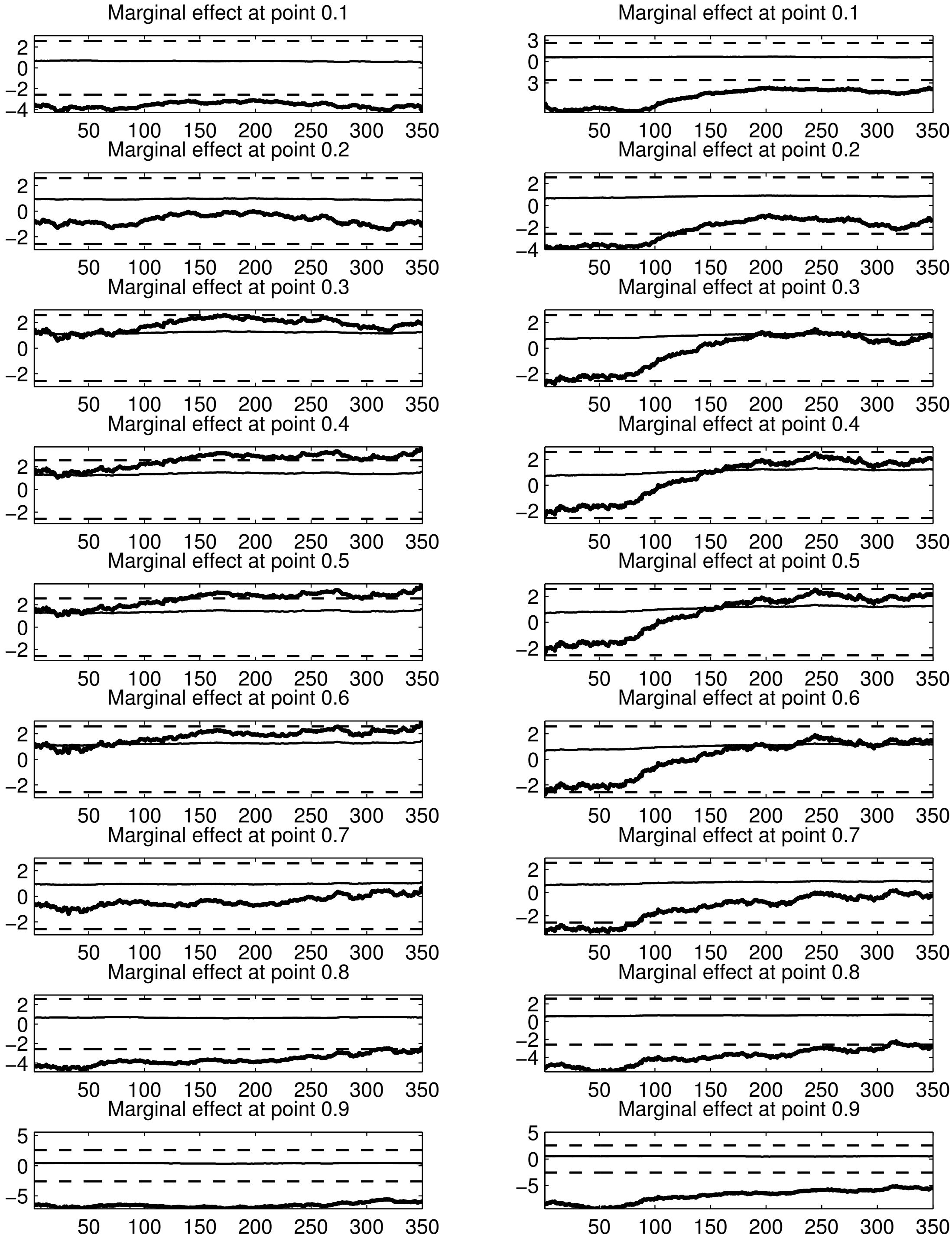
The estimation results are graphically presented on Figure 3. Market and simulated prices share some common features. First, both prices react correctly in the range of $0.2 - 0.3$ and $0.6 - 0.7$. Second, there is the effect of overreaction for extreme prices of 0.1 and $0.8 - 0.9$. One explanation for the overreaction of market prices could be the lower constraint on the tick size: any desired change below the tick size will result in overreaction. But since this phenomenon is common and simulated prices are not constrained, the other explanation is more likely: extreme events are not so frequent in the sample of the DJIA index and also rarely observed by market participants, therefore it is difficult to derive their correct probabilities. Third, underreaction of the prediction market prices in the range of $0.4 - 0.5$ is their unique feature. In this range a tiny change in moneyness transforms into a big change in the price, therefore, traders are not sensitive enough. It might be explained by the fact that the market is not deep enough, therefore it could be possible to change the price only gradually.

Our next experiment will be to run the following probit regression (regression-based test of the information content following Fair and Shiller, 1990):

$$\mathbb{E}(P_{j,T} = 1 | P^1_{j,t}, P^2_{j,t}) = \Phi(\beta_{0,t} + \beta_{1,t} P^1_{j,t} + \beta_{2,t} \frac{P^2_{j,t}}{P^1_{j,t}}), \tag{7}$$

where $P^1_{j,t}$ and $P^2_{j,t}$ stay for market and simulated prices in one regression and in the reversed order in the other (I add a small constant to simulated prices, when they are in the denominator, because some values are exactly zero). The ratio stays for the difference between two predictions and helps to avoid potential multicollinearity in the case of regressing simply on $P^2_{j,t}$. Then one could consider the significance of the coefficient β_2 in both regressions. If the coefficient is insignificant, it means that the additional regressor does not bring any new information. The graphs clearly show us that the market prices mostly contain additional information compared to the simulated prices, but rarely vice versa.

Figure 3: Marginal effect of change in market and simulated prices



Note: The graphs show the dynamics of the marginal effect (2) of change in the price (probability) for the market price (left column) and the simulated price (right column). Marginal effects are evaluated at every trading minute (from the 1st to the 350th minute) and for the values of probability $\bar{P}_{j,t} = 0.1, 0.2, \ldots, 0.9$. Marginal effects are shown with the thin line, statistics τ (6) with the thick line, and the lower and upper borders of the 99% confidence interval with the dashed line.

Another way to compare the predictive ability of market and simulated prices is to use analogues of the measure of goodness of fit R^2 developed for binary regressions. I use two measures: Pseudo R^2 of McFadden (McFadden, 1974) and Brier Score (Brier, 1950). McFadden's Pseudo R^2 is equal to one minus the ratio of the log likelihood of the regression with all the predictors included (unrestricted) to the log likelihood of the regression using a constant as the only predictor (the other coefficients are restricted to be zero). Since the log likelihood in this case is always negative and does not decrease after adding other predictors to the constant, McFadden's Pseudo R^2 lies between 0 and 1 and a higher value of this measure corresponds to a better fitting model.

$$R^2_{McFadden,t} = 1 - \frac{\log \mathcal{L}_{UR}(\hat{\beta}_t)}{\log \mathcal{L}_R(\tilde{\beta}_t)}, \quad \hat{\beta}_t = \text{argmax}\{\mathcal{L}_{UR}(\beta_t)\}, \quad \tilde{\beta}_t = \text{argmax}\{\mathcal{L}_R(\beta_t)\} \tag{8}$$

$$\log \mathcal{L}_{UR}(\beta_t) = \sum_{j=1}^{J} \{P_{j,T} \log \Phi(\beta_{0,t} + \beta_{1,t} P_{j,t}) + (1 - P_{j,T}) \log(1 - \Phi(\beta_{0,t} + \beta_{1,t} P_{j,t}))\} \tag{9}$$

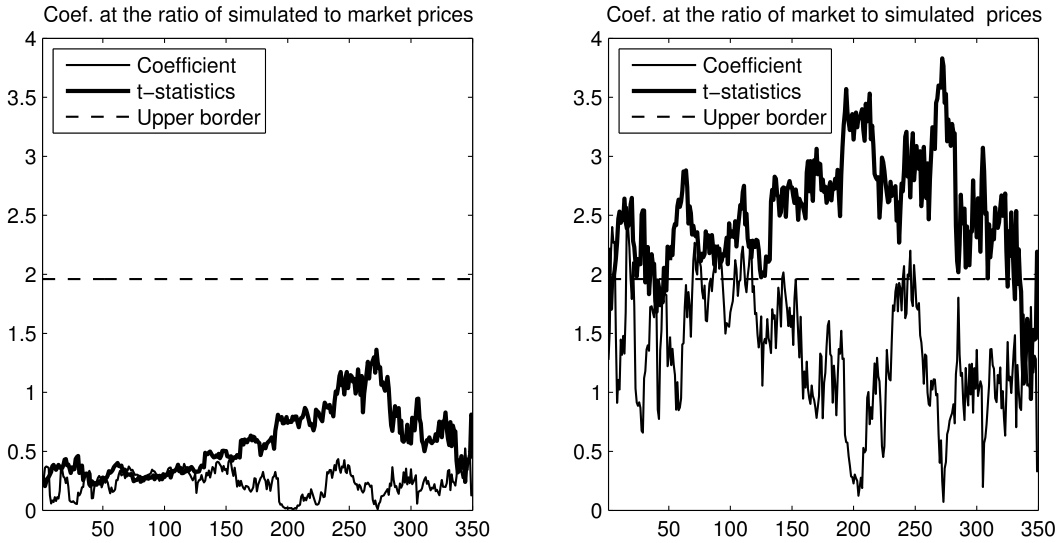
$$\log \mathcal{L}_R(\beta_t) = \log \Phi(\beta_{0,t}) \sum_{j=1}^{J} P_{j,T} + \log(1 - \Phi(\beta_{0,t})) \sum_{j=1}^{J} (1 - P_{j,T}) \tag{10}$$

The Brier score is equal to the mean squared difference between the actual and predicted outcomes and is in our case model-free, because price is by itself a prediction. The lower is the Brier score, the lower is the discrepancy between the prediction and the outcome.

$$BrierScore_t = \frac{\sum_{j=1}^{J} (P_{j,T} - P_{j,t})^2}{J} \tag{11}$$

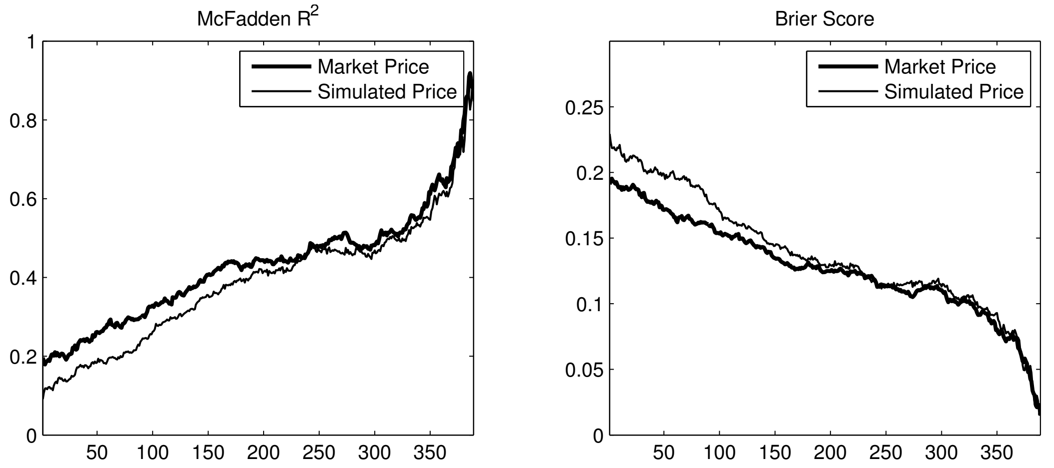
Figure 5 shows that as the expiration time comes closer, the performance of two prices converges, but in the first half of the day the predictive ability of the market price is better than that of the simulated price.

Figure 4: Test for the information content



Note: Coefficient $\beta_{2,t}$ (thin line) and corresponding t-statistics (thick line) for the regression $\mathbb{E}(P_{j,T} = 1|P_{j,t}^1, P_{j,t}^2) = \Phi(\beta_{0,t} + \beta_{1,t}P_{j,t}^1 + \beta_{2,t}\frac{P_{j,t}^2}{P_{j,t}^1})$. $P_{j,t}^1$ and $P_{j,t}^2$ stay for market and simulated prices in the regression on the left graph and in the reversed order in the regression on the right graph. On the X-axis are minutes.

Figure 5: Predictive ability comparison



Note: McFadden R^2 (left) and Brier score (right) at each moment t for regressions on market (thick line) and simulated (thin line) prices. On the X-axis are minutes.

5 Pricing Function

The most important determinants of the option price suggested in the literature are moneyness, the time to maturity and the implied volatility. In many cases the implied volatility is modeled either from the data with a higher frequency (which I do not have), or from the process driving the dynamics of the underlying asset under a model framework. I would like to avoid making additional assumptions and defending a model, therefore, I treat the price of the option as a function only from the time to maturity and moneyness. The moneyness on day j at time t is again defined as

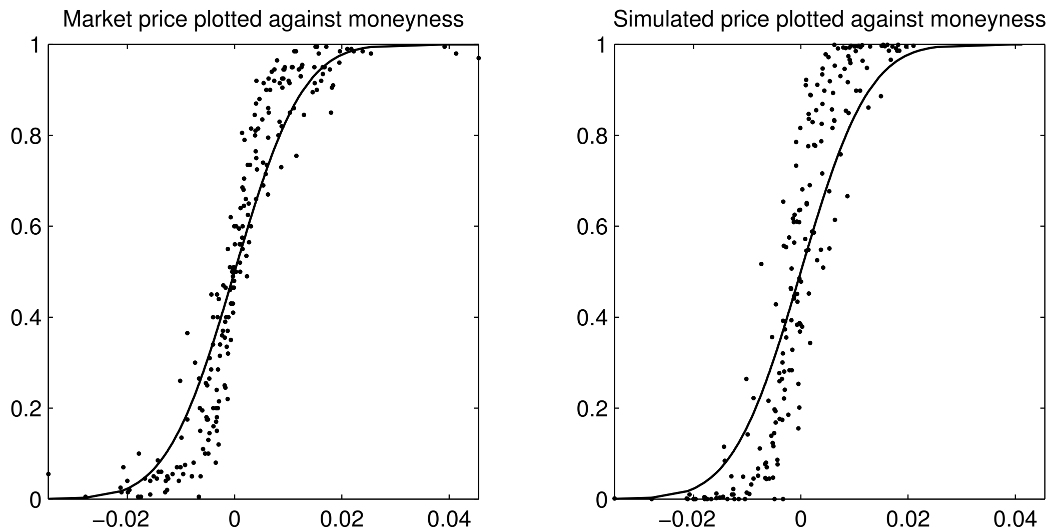
$$M_{j,t} = \log S_{j,t} - \log C_{j-1}, \tag{12}$$

where C_{j-1} denotes the close price of the previous day $j-1$ and $S_{j,t}$ is the price of the DJIA index on day j at time t. I ignore discounting while deriving approximations for the prices, since the intraday risk-free rate is very close to zero.

Let $f_t(x)$ be the density of $u_t = \log S_{j,T} - \log S_{j,t}$, the DJIA index return during the time span $[t,\, T]$, and $F_t(x)$ be its cumulative distribution function. Then the price of the option on day j at time t is

$$P_{j,t}(M_{j,t}) = \int_{-M_{j,t}}^{\infty} f_t(u_t)\, du_t = 1 - F_t(-M_{j,t}), \tag{13}$$

Figure 6: Market and simulated prices as functions of moneyness, the 195th trading minute

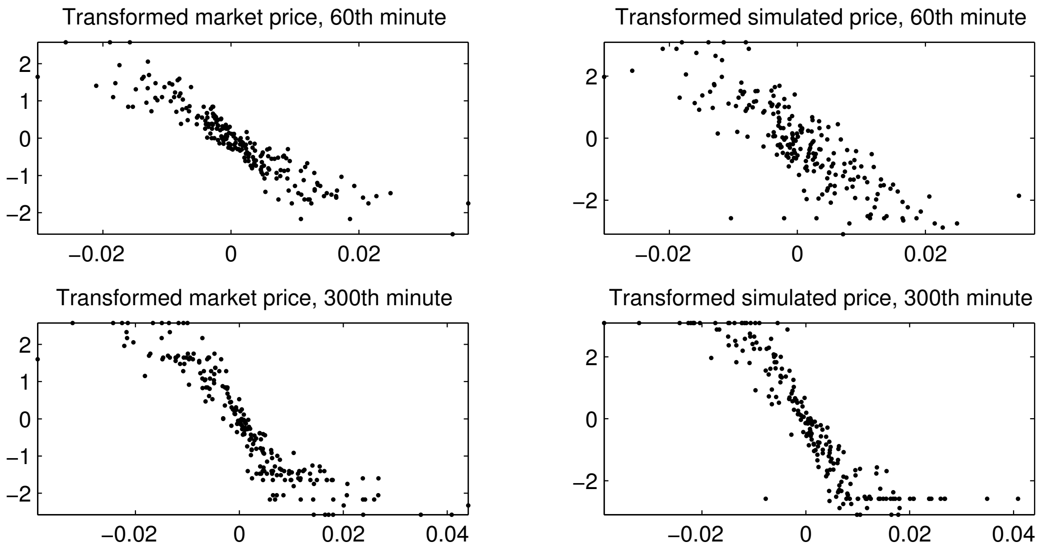


Note: Market and simulated prices of digital options on the DJIA index in the middle of the trading day (195th minute) plotted against corresponding moneyness. The prices are normalized to one. Function $\Phi(100 \cdot Moneyness; 0, 1)$ is shown for comparison, where $\Phi(x; 0, 1)$ is the cumulative distribution function of the standard normal distribution.

If one takes a look at Figure 6, one can easily notice that the shape of the empirical mapping of moneyness to prices resembles the shape of the function $\Phi(x, 0, 1)$, where $\Phi(x, 0, 1)$ is the cumulative distribution function of the standard normal distribution. Let us take this observation into account and apply the following transformation to the option prices:

$$\tilde{P}_{j,t} = \Phi^{-1}(P_{j,t}; 0, 1). \tag{14}$$

Figure 7: Transformed prices as functions of time and moneyness



Note: Transformed market and simulated prices at the 60th and the 300th trading minutes plotted against corresponding moneyness. The transformation is defined in (14). The prices are normalized to one. The values on the X-axis correspond to moneyness.

It could be seen from Figure 7 that the dependence between the transformed prices and moneyness is in general linear, with some non-linearities at the edges of the plot. Also the "slope" of the mapping seems to change during the day. Therefore, I suggest the following polynomial $g(t, M)$ to fit the transformed prices (recall that t denotes time (in minutes) and M stays for moneyness):

$$g(t, M) = \sum_{k=0}^{3} (\alpha_k + \beta_k \sqrt{t}) M^k. \tag{15}$$

The estimates of the parameters for market and simulated prices are given in Table 7.

Table 7: Estimates of polynomial coefficients

	α_k		β_k	
	Market Price	Simulated Price	Market Price	Simulated Price
$k=0$	0.0038	-0.2462	0.0011	0.0114
	(0.70)	***(-26.59)	**(2.76)	***(17.18)
$k=1$	-91.4503	-107.5778	-2.8950	-5.0404
	***(-129.45)	***(-87.57)	***(-61.47)	***(-62.13)
$k=2$	114.9480	509.6294	-17.2159	-20.2549
	***(3.82)	***(9.65)	***(-8.71)	***(-5.87)
$k=3$	63909.4223	83507.4581	-817.3293	-692.9024
	***(56.63)	***(38.18)	***(-11.85)	***(-5.30)

Note: The table presents the coefficients of the polynomial (15) fitted to market and simulated prices of the digital options. t-statistics are given in parentheses. Levels of significance are the following: * is 5%, ** is 1% and *** is 0.1%.

The impact of a small change in moneyness on the option price evaluated at the point $(\bar{t}, \bar{M})$ is

$$\frac{\partial P(\bar{t}, \bar{M})}{\partial M} = \phi\left(\sum_{k=0}^{3}(\alpha_k + \beta_k\sqrt{\bar{t}})\bar{M}^k\right)\sum_{k=1}^{3} k\left(\alpha_k + \beta_k\sqrt{\bar{t}}\right)\bar{M}^{k-1} \tag{16}$$

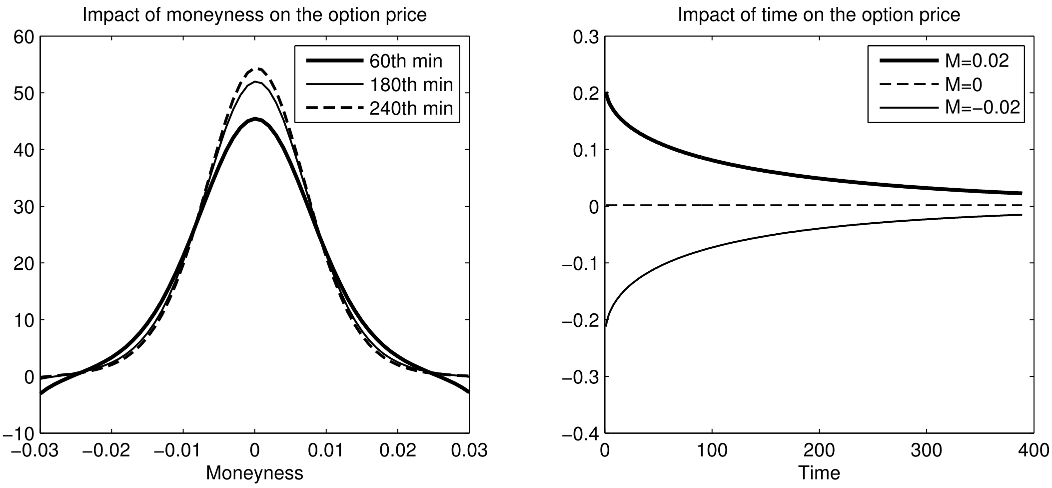
and the impact of a small change in the time to maturity is

$$\frac{\partial P(\bar{t}, \bar{M})}{\partial t} = \phi\left(\sum_{k=0}^{3}(\alpha_k + \beta_k\sqrt{\bar{t}})\bar{M}^k\right)\sum_{k=1}^{3}\left(\alpha_k + \beta_k\frac{1}{2\sqrt{\bar{t}}}\right)\bar{M}^k. \tag{17}$$

The left side of Figure 8 shows the magnitude of the impact of a change in moneyness for different $\bar{M}$ and three different moments of the day. The impact for relatively large absolute values of moneyness almost does not change during the day, but the effect of moneyness for at-the-money options grows. The impact of time is simple: for a given negative moneyness the price goes down with time and for a given positive moneyness the price goes up.

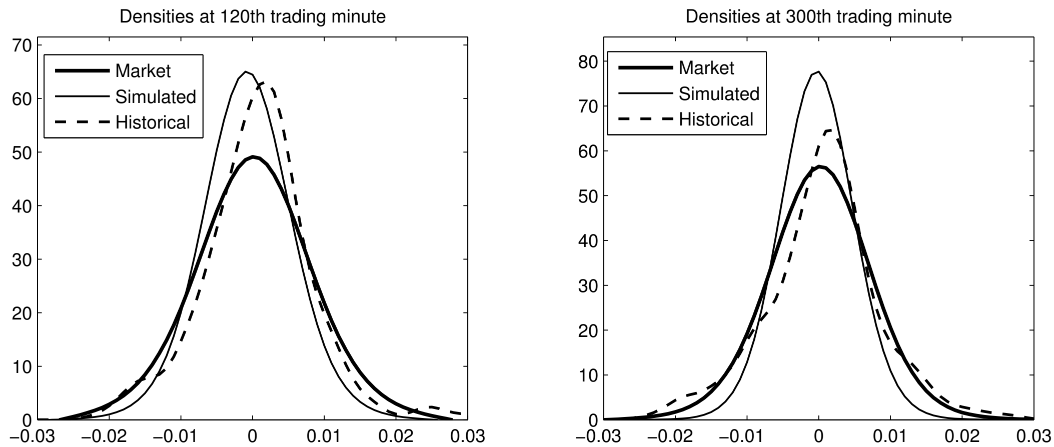
The found polynomial approximation allows us also to approximate the density function of $u_t = \log S_{j,T} - \log S_{j,t}$ at each moment t: it follows from (13) that $f_{t,u_t}(\bar{M}) = \frac{\partial P(t, -\bar{M})}{\partial M_{j,t}}$. Figure 9 shows three densities at the 120th and the 300th trading minute. The first density is derived from the market prices, the second density from the simulated prices and the third density is non-parametrically estimated from the subsamples of $\log S_{j,T} - \log S_{j,t}$, where j runs through all days in the sample. One can easily notice that the historical density is positively skewed, so it clearly underestimates the risk. The other two densities are only slightly positively skewed. Remarkable is that the density derived from the simulated prices has much thinner tails than the density derived from the market prices. It means that if the market prices are indeed closer to the true probability than the simulated prices, they do contain additional information about probabilities of extreme events.

Figure 8: Impact of moneyness and time on the option prices



Note: The impact of a change in moneyness (16) for different $\bar{M}$ and three different moments of the trading day (left) and the impact of a change in time (17) for different $\bar{t}$ and three different values of moneyness (right).

Figure 9: Estimated density functions



Note: The density of $\log S_{j,T} - \log S_{j,t}$ derived in three different ways in the second (left) and the last (right) hour of the day. The dashed lines correspond to the densities fitted non-parametrically to u_t calculated from the DJIA index values. The solid lines correspond to the densities derived using the suggested polynomial approximation.

6 Trading Strategies

As the price of the option is a function of the price of its underlying asset, it is clear that there must be a detectable dependency between the option price and contemporaneous and lagged characteristics of the DJIA index, such as returns and moneyness. These dependencies have been examined in the section discussing the microstructure of the prediction market. In this section I test a bold assumption that a new equilibrium market price consists not only of an adjustment to the realized changes of economic environment, but also of a prediction of some future changes. This could be one of the reasons why the option price does not always react to the fluctuations of the DJIA index: this information has been already taken into account by the prediction market and is incorporated into the option price.

Certainly, one should not expect that the prediction market can predict returns on the Dow Jones market, because the latter is incomparably deeper than the former. But I anticipate to be able to predict at least directions of the DJIA index changes.

To test this assumption, I run a probit regression, in which the dependent variable is a binary indicator of an upward movement of the DJIA index within a certain time interval. Time intervals i vary from 5 minutes to 1 hour. The explanatory variables are changes in the option price in the current and previous trades $\Delta P_{j,\tau}$ and $\Delta P_{j,\tau-1}$, current moneyness $M_{j,\tau}$ (not multiplied by 1000 here) and a trading moment τ. The specification with the traded volume was also tested, but the volume does not seem to have any predictive power.

Table 8 reports the results. It could be noticed that for short horizons up to 15 minutes the contemporaneous and lagged option price changes and moneyness demonstrate some explanatory power for the direction of the DJIA index movements, for longer horizons only coefficients of time and the current option price change are significant, and for the time interval of 1 hour the only predictor with the significant coefficient is a constant. In this specification changes in the option price stay for the momentum effect, because an upward movement of the option price increases the probability of an upward movement of the DJIA index. Moneyness stays for the reversal effect: if moneyness decreases, meaning that the DJIA index has gone down, then the probability of its subsequent upward movement increases. But this effect is not very prominent, since it is significant only for 10 and 15 minute horizons.

Table 8: Predictability of the DJIA index upward movements for different horizons

	5 min	10 min	15 min	20 min	25 min	30 min	45 min	60 min
$Const$	0.0315	0.0232	0.0362	0.0175	0.0301	0.0130	0.0555	0.0792
	(1.30)	(0.95)	(1.47)	(0.70)	(1.21)	(0.52)	*(2.15)	**(2.99)
$\Delta P_{j,\tau}$	2.0556	1.5634	1.3937	1.2843	1.0906	0.8235	0.7865	0.2476
	***(9.39)	***(6.93)	***(5.97)	***(5.28)	***(4.41)	**(3.26)	**(2.95)	(0.90)
$\Delta P_{j,\tau-1}$	0.5781	0.5272	0.4572	0.3592	0.1352	0.1023	0.3157	-0.1660
	**(2.67)	*(2.35)	*(1.97)	(1.48)	(0.55)	(0.41)	(1.19)	(-0.60)
$M_{j,\tau}$	-3.4211	-4.2489	-4.8030	-2.0314	-1.8966	0.4606	-0.0142	0.4331
	(-1.76)	*(-2.16)	*(-2.42)	(-1.01)	(-0.94)	(0.23)	(-0.01)	(0.20)
τ	-0.0001	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004	0.0001
	(-0.68)	(1.63)	(1.54)	**(2.69)	**(2.67)	**(3.16)	**(3.05)	(0.62)
N of obs.	9946	9659	9350	9065	8854	8625	7931	7345

Note: The specification of the probit regression is the following. $\mathbb{E}(\text{sgn}\,\Delta\log S_{j,\tau+i} = 1|X) = \Phi(\beta X)$, $\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$. Time intervals i vary from 5 minutes to 1 hour. The regressors are changes in the option price in the current and previous trades $\Delta P_{j,\tau}$ and $\Delta P_{j,\tau-1}$, current moneyness $M_{j,\tau}$ and a trading moment τ. $\{j,\tau\}$ is the pooled index of an observation. t-statistics are given in parentheses. Levels of significance are the following: * is 5%, ** is 1% and *** is 0.1%.

The next step is to find a practical application for the detected dependencies. Since I consider the DJIA index only, one of the possible strategies, where the prediction market prices could be applied, is the following: buy (sell) one unit of the DJIA index today, sell (buy) this unit in i minutes and fix the profit (loss). The question will be then only how to decide whether to trade and if to trade than to buy or to sell.

I suggest the following simple decision rule: each trading minute, when there is an activity in the prediction market and, therefore, a new information arrival, go through the steps listed below.

1. Estimate the regression with the specification given in Table 8 using a historical data with a fixed window length (in my example the length is set to 60 days (23400 trading minutes), which roughly corresponds to a quarter of the sample).

2. For each new arriving trade forecast the probability of an upward movement of the Index in the next i minutes ($i = 5, 10, 15$) using the coefficient estimates of the regression.

3. Using two ex ante chosen thresholds, upper and lower, make a decision about the current action, comparing the probability and the thresholds. If the probability lies between the thresholds, one should not take any action. If the probability is above the upper threshold, one should buy the Index. If the probability is below the lower threshold, one should sell the Index. In i minutes the Index is bought/sold back and the profit/loss is fixed.

I consider all possible combinations of upper and lower thresholds on the grid with the size of 0.01 and in order to be able to compare the results with other decision rules later, I report only the combinations of thresholds, which lead to approximately 1000 trades.[2] I consider the first three forecast intervals, where two or three coefficients of the important explanatory variables are significant. Earnings per trade stay for the average sum of money in dollars earned in a trade, the average return and the Sharpe ratio are calculated on the sample of the Index trades only.[3]

Table 9 shows the results. It is noticeable that the performance of the strategy worsens with the distance of prediction. The best results are achieved for the values of the upper threshold around 0.57 and these of the lower threshold around 0.45. However, these numbers cannot tell us, whether the strategy is good or not, we need to compare it with some other strategies.

Table 9: Trading strategies based on the prediction market information

Upper Threshold	Lower Threshold	Number of Trades	Earnings per Trade	Average Return	Sharpe Ratio
5 min interval					
0.75	0.46	1063	2.8063	0.0003	0.2430
0.62	0.45	981	2.9928	0.0003	0.2655
0.59	0.44	972	2.6652	0.0003	0.2343
0.57	0.43	1047	**3.0139**	0.0003	**0.2660**
0.56	0.41	1019	2.8501	0.0003	0.2449
0.55	0.35	1004	2.0163	0.0002	0.1784
10 min interval					
0.6	0.48	1025	2.4494	0.0003	0.1472
0.57	0.47	1055	2.4416	0.0003	0.1528
0.56	0.45	973	**2.5314**	0.0003	**0.1672**
0.55	0.35	1071	1.4380	0.0002	0.1082
0.54	0.25	1065	1.4574	0.0002	0.1094
15 min interval					
0.61	0.49	987	2.0952	0.0002	0.1071
0.58	0.48	1033	2.0983	0.0002	0.1166
0.57	0.46	993	**2.1673**	0.0002	**0.1300**

Note: The table shows financial results of the strategy based on the information from the prediction market. The combinations of thresholds are set so that the number of trades is approximately 1000. Three time intervals for prediction of the probability of an upward movement of the Index are considered.

Since the number of possible decision rules is uncountable, it could be possible to analyze by simulations what average trade earnings the best of them provide. To do so, I vary the expected

[2]I have repeated all the calculations fixing the target number of trades to be 500, and all conclusions have remained the same. These results are not reported for the sake of brevity.

[3]Moments of no trading activity with zero return are not considered, but since the number of trades is kept roughly the same, the comparison is correct.

proportion of buying in trades (from 0.1 to 0.9) and fix the expected number of trades to be 1000.

Table 10: Random trading strategies

	95% Quantile			99.9% Quantile			Maximum		
Share of Buyings	Earnings per Trade	Average Return	Sharpe Ratio	Earnings per Trade	Average Return	Sharpe Ratio	Earnings per Trade	Average Return	Sharpe Ratio
5 min interval									
0.7	0.5147	0.0001	0.0547	0.9345	0.0001	0.1000	1.2709	**0.0002**	**0.1367**
0.8	0.5377	0.0001	0.0570	**0.9726**	0.0001	**0.1102**	**1.3083**	0.0001	0.1336
0.9	**0.5590**	0.0001	**0.0593**	0.9690	0.0001	0.1024	1.1791	0.0001	0.1140
10 min interval									
0.8	0.7357	0.0001	0.0579	1.3025	0.0001	**0.1053**	1.5550	0.0002	0.1332
0.9	**0.7413**	0.0001	**0.0587**	**1.3162**	0.0001	0.1040	**1.7185**	0.0002	**0.1359**
15 min interval									
0.8	0.9011	0.0001	0.0590	**1.6036**	0.0002	0.1004	**2.2033**	0.0002	0.1309
0.9	**0.9158**	0.0001	**0.0605**	1.5777	0.0002	**0.1050**	1.7463	0.0002	**0.1380**

Note: The table presents the best results of random buying and selling with different proportions of buying and the fixed number of trades of 1000. The number of simulated strategies is 10000. Only rows containing the highest value in a column are reported to preserve space.

Table 10 reports the maximum and the 99.9% and 95% quantiles of the financial characteristics of 10000 simulated random strategies with different buying and trading proportions and for three different time intervals (only rows containing the highest value in a column are reported to preserve space). We see that higher average earnings are provided by a lower proportion of selling (0.1-0.2). One can notice that, first, for the 5 minute interval the corresponding average trade earnings of the prediction market based strategy are always greater than the maximum earnings of random strategies. As the time interval increases, the predictive performance of the information market worsens, but random strategies can beat it only for the 15 minute interval of forecasting.

Certainly, the most popular trading rules are not random, but rely on some signals from the market. Sullivan et al., 1999 describe six popular technical strategies, which could result in a thousand of trading rules depending on parameters. Three of these six strategies could be applied to my setting: Filter Rule, Moving Averages and Supply & Resistance. Each of these rules is considered for three horizons and with such parameters that the resulting number of trades is close to 1000.

I define the Filter Rule strategy with parameters U and D as follows: If the current value of the DJIA index moves up at least U basis points in comparison to the price 15 minutes ago, then buy the Index. If the current value moves down at least D basis points in comparison to the price 15 minutes ago, then sell the Index. Distance of 15 minutes is chosen because it provides the highest earnings and the Sharpe ratios.

Table 11: Filter Rule trading strategies

Parameter U, bp	Parameter D, bp	Number of Trades	Earning per Trade	Average Return	Sharpe Ratio
5 min interval					
66	-39	998	0.6265	0.0001	0.0422
59	-40	996	0.8837	0.0001	0.0611
45	-45	1003	1.2757	0.0001	0.0914
42	-49	998	1.5871	0.0002	0.1160
39	-55	1002	1.8851	0.0002	0.1401
38	-58	1005	**2.1178**	0.0002	0.1603
37	-64	999	1.9144	0.0002	0.1492
36	-73	999	2.0474	0.0002	**0.1684**
10 min interval					
76	-38	1000	1.4275	0.0002	0.0706
63	-39	1000	2.2171	0.0002	0.1107
52	-41	1000	2.8210	0.0003	0.1462
47	-43	998	3.1601	0.0003	0.1689
41	-50	999	3.8198	0.0004	0.2127
37	-62	1001	**4.1293**	0.0005	0.2409
36	-70	1000	4.1162	0.0005	**0.2500**
15 min interval					
70	-38	1000	2.3350	0.0002	0.0966
46	-43	1002	4.2889	0.0005	0.1849
45	-44	998	3.9752	0.0004	0.1708
41	-49	1000	4.4959	0.0005	0.2001
38	-56	1000	4.5379	0.0005	0.2109
36	-66	998	**4.7587**	0.0005	**0.2307**
35	-85	1000	4.1820	0.0005	0.2156

Note: The table presents the results of Filter Rule trading strategies with different values of U and D parameters. The pairs of parameters are set so that the number of trades is approximately 1000. Three time intervals for prediction of the probability of an upward movement of the Index are considered.

Table 11 shows financial results of the Filter Rule trading rules. The overall performance of this strategy is excellent. However, for the 5 minute horizon the prediction market based trading rules mostly outperform the Filter Rule. Unfortunately, for 10 and 15 minute horizons there are many parameter combinations for which the Filter Rule trading rules show much greater earnings per trade, returns and the Sharpe ratios than the prediction market based trading rules.

The Moving Averages strategy with parameters F and S, $F < S$ could be described in such a way: If the average value of the DJIA index over the last F minutes becomes greater than the average value over the last S minutes, then sell the Index. If the average value over the last F minutes becomes lower than the average value over the last S minutes, then buy the Index. Commonly, under the first condition one should buy and under the second sell, but it turns out that the DJIA index with one-minute frequency displays strong mean-reversion, and one should redefine the rule to get positive returns.

Table 12: Moving Averages trading strategies

Parameter F, min	Parameter S, min	Number of Trades	Earning per Trade	Average Return	Sharpe Ratio
5 min interval					
4	209	1002	0.4343	0.0000	0.0505
5	190	1000	0.5072	0.0001	0.0580
6	164	998	**0.6718**	0.0001	**0.0792**
10	120	998	0.5922	0.0001	0.0736
14	97	1001	0.3918	0.0000	0.0474
15	94	1000	0.3273	0.0000	0.0420
16	88	998	0.1921	0.0000	0.0254
19	82	1000	0.2184	0.0000	0.0240
26	74	1001	0.0468	0.0000	0.0072
10 min interval					
2	299	1000	0.2862	0.0000	0.0202
3	252	999	0.4591	0.0001	0.0341
5	181	1002	0.4550	0.0000	0.0333
10	115	1002	0.3771	0.0000	0.0296
14	96	1001	**0.5101**	0.0001	**0.0427**
15	91	1001	0.3597	0.0000	0.0315
17	86	1002	0.3675	0.0000	0.0301
18	82	1002	0.2707	0.0000	0.0207
19	78	1000	0.4278	0.0000	0.0335
26	73	998	0.1180	0.0000	0.0104
15 min interval					
3	221	1001	-0.1731	-0.0000	-0.0072
4	199	998	0.0632	0.0000	0.0074
8	126	1002	0.5054	0.0001	0.0319
9	121	1001	0.4819	0.0001	0.0328
10	114	998	**0.6672**	0.0001	**0.0449**
18	79	1001	0.3523	0.0000	0.0221
24	72	999	0.1447	0.0000	0.0108
28	71	998	0.2968	0.0000	0.0207

Note: The table presents the results of Moving Averages trading strategies with different values of F and S parameters. The pairs of parameters are set so that the number of trades is approximately 1000. Three time intervals for prediction of the probability of an upward movement of the Index are considered.

Table 12 presents the results, which are not impressive and are always outperformed by the prediction market based strategies.

The last considered strategy is Support & Resistance with parameters H and L. If the current value of the DJIA index is greater than the maximum value over the last H minutes by at least 1 basis point, then buy the Index. If the current value is lower then the minimum value over the last L minutes by at least 1 basis point, then sell the Index. The threshold of 1 basis point is chosen because it provides the required number of trades.

Table 13: Support & Resistance trading strategies

Parameter H, min	Parameter L, min	Number of Trades	Earning per Trade	Average Return	Sharpe Ratio
5 min interval					
5	99	1000	0.6318	0.0001	0.0452
6	49	1000	0.5636	0.0001	0.0395
11	20	1000	0.7484	0.0001	0.0543
54	7	1000	0.7781	0.0001	**0.0565**
80	6	1000	**0.7786**	0.0001	0.0561
189	5	1000	0.7227	0.0001	0.0512
362	4	1000	0.5066	0.0001	0.0362
10 min interval					
4	199	1000	0.7617	0.0001	0.0414
7	34	1000	0.9617	0.0001	0.0496
26	10	1000	1.1250	0.0001	0.0612
31	9	1000	1.0860	0.0001	0.0594
35	8	1000	1.0963	0.0001	0.0603
168	5	1000	**1.3082**	0.0001	**0.0690**
325	4	1000	1.1788	0.0001	0.0621
15 min interval					
3	338	1000	1.5850	0.0002	0.0692
4	131	1000	1.4048	0.0002	0.0596
5	55	1000	**1.9968**	0.0002	**0.0815**
6	34	1000	1.7726	0.0002	0.0715
11	15	1000	1.3940	0.0001	0.0560
16	11	1000	1.4579	0.0002	0.0591
102	5	1000	1.5270	0.0002	0.0629
265	4	1000	1.0188	0.0001	0.0423

Note: The table presents the results of Support & Resistance trading strategies with different values of H and L parameters. The pairs of parameters are set so that the number of trades is approximately 1000. Three time intervals for prediction of the probability of an upward movement of the Index are considered.

The results of the application of the Support & Resistance trading strategies are in Table 13. On average, this strategy performs better than the Moving Averages trading strategies, but worse than the Filter Rule trading strategies. Only for the 15 minute forecasting horizon, the performance is approaching that of the prediction market based strategies in terms of earnings per trade and average returns, but the Sharpe ratios remain much lower.

To draw a conclusion, for the 5 minute forecasting horizon the performance of the prediction market based strategies is the best compared to 10000 random strategies with 9 different buying-selling proportions and to three widely used trading rules. As the horizon increases, the performance deteriorates but still beats a great number of alternative trading rules.

7 Conclusion

This paper considers the online trading platform for prediction markets Intrade.com, the data sources of which have not been widely used in the literature yet. I examine in detail the microstructure of the market for binary options on the Dow Jones Industrial Average index. Its prominent feature is the availability of the measurable driving process as opposed to e.g. a political prediction market where the driving process is unobserved.

I show that the prediction market prices are closer to the true unobservable probabilities because they better predict the final outcome than the virtual probabilities derived from the the DJIA index values only. They also contain some additional information not incorporated in the Index. Motivated by this finding, I develop a simple trading strategy based on the trading activity on the prediction market and compare its performance to a large number of other possible strategies. Very good financial results of the prediction market based strategies for a short forecasting horizon increase the optimism about its potential practical application.

The further research in this area could go in two directions. First, the suggested trading strategy might be improved so that it outperforms all considered alternative trading rules presented in this paper. Its performance is worth testing on the DJIA futures. To improve the strategy, one can try to use the data from other binary options on the DJIA index, e.g. these expiring at 100 points if the Index exceeds/falls behind the previous close by a particular number of points. Second, one might develop a trading strategy for the prediction market itself, because many researchers believe that the trading volume and trading activity on these markets will grow in the future.

References

[1] N. Archak and P. Ipeirotis. Modeling volatility in prediction markets. Working Paper CEDER-08-07, 2008.

[2] J. Berg, F. Nelson, and Rietz T. Prediction market accuracy in the long run. *International Journal of Forecasting*, 24:285–300, 2008.

[3] T. Berry and K. Howe. Public information arrival. *The Journal of Finance*, 49:1331–1346, Sep. 1994.

[4] G. Brier. Verification of weather forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

[5] R. Fair and R. Shiller. Comparing information in forecasts from econometric models. *American Economic Review*, 80(3):375–389, 1990.

[6] W. H. Greene. *Econometric Analysis*. Prentice-Hall, 5 edition, 2003.

[7] R. Gürkaynak and J. Wolfers. Analysis of market-based macro forecasts, uncertainty, and risk. In *NBER International Seminar on Macroeconomics, edited by C. Pissarides and J. Frankel*. Cambridge, MIT Press, 2006.

[8] S.N. Lahiri. *Resampling methods for dependent data*. Springer, New York, 2003.

[9] Ch. Lee and M. Ready. Inferring trade direction from intraday data. *The Journal of Finance*, 46(2):733–746, June 1991.

[10] D. McFadden. *Frontiers in Econometrics*, chapter Conditional logit analysis of qualitative choice behavior, pages 105–142. Academic Press: New York, 1974.

[11] M. Ottaviani and P. Sørensen. The favorite-longshot bias: An overview of the main explanations. In *Handbook of Sports and Lottery Markets, edited by D. Hausch and W. Ziemba*, pages 83–101. North-Holland, Amsterdam, 2008.

[12] D. Pennock and M. Wellman. Representing aggregate belief through the competitive equilibrium of a securities market. In *Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 392–340, San Francisco, CA, 1997.

[13] R. Sullivan, A. Timmermann, and H. White. Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance*, 54:1647–1691, 1999.

[14] P. C. Tetlock. Liquidity and prediction market efficiency. Working Paper, 2008.

[15] G. Tziralis and I. Tatsiopoulos. Prediction markets: an extended literature review. *The Journal of Prediction Markets*, 1:75–91, 2007.

[16] J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, Spring 2004.

[17] J. Wolfers and E. Zitzewitz. Interpreting prediction market prices as probabilities. NBER Working Paper No. 12200, 2006.

[18] E. Zitzewitz. Price discovery among the punters: Using new financial betting markets to predict intraday volatility. July 2006.

Portfolio Delegation and Fiduciary Management

This paper was presented at

- Research Seminar of the Department of Banking & Finance, Zurich, Switzerland (November 2011)
- 39th Annual Meeting of the European Financial Association, Copenhagen, Denmark (August 2012)

Abstract

The fiduciary manager is a third party between an asset owner and an asset manager, delegated to improve their communication and mitigate problems of information asymmetry. The information available to each of these parties differs in its amount and precision. This paper studies the impact of the quality of agents' information on the investor's decision about portfolio delegation. If the fiduciary managers' information is precise, hiring them is always more beneficial for investors than delegating the portfolio management to an asset manager. An investor, who works with a fiduciary manager possessing imperfect information, still achieves some gains if surrounded by badly informed asset managers. An investigation of the general equilibrium shows that the prices of assets included in portfolios of asset managers rise with the expansion of fiduciary management and the prices of assets which are left aside decrease.

Keywords: Portfolio Delegation; Decentralized Management; Fiduciary Management; Advisory Services; Asymmetric Information

JEL classification: G11; G20; D82

1 Introduction

The concept of fiduciary management was developed in the Netherlands in the beginning of the 21st century. The fiduciary manager is a third party between an asset owner and an asset manager, who is delegated to improve their communication and mitigate the problems of information asymmetry. This practice has recently become an important role in the management structure of pension funds and insurance companies. According to the article in Professional Pensions (26 Feb 2010), there are several key reasons why fiduciary management is expanding. First, insufficient time consistency of investment strategies and a lack of resources to deal with their increasing complexities; second, more demanding regulation and reporting requirements; third, the importance of manager selection. The comment of Goldman Sachs Asset Management (2005) summarizes the main duties of the fiduciary management: strategic investment management and advisory; efficient portfolio building and risk budgeting to meet the long-term investment objectives; portfolio manager selection; investment performance monitoring; customized and consolidated reporting.

This study is mainly focusing on two aspects of the fiduciary management: asset manager selection and efficient portfolio building. I develop a three-tier model of the interaction between the principal (an asset owner or an investor), the intermediary (a fiduciary manager) and many agents (portfolio managers). The economy participants share some common information about the economic environment and possess some private information. The private information is not shared or sold, but can be used to improve the economic welfare if the interaction is properly organized. In this framework I examine benefits and losses of using the services of the fiduciary manager for the investor in terms of her utility function. Further I characterize the circumstances, in which such services are profitable for the investor. I also analyse the new economic equilibrium that arises after establishing a fiduciary management institution.

The key actors of the model are investors, fiduciary managers and asset managers. The investors possess wealth, but their ability to allocate it is restricted. The asset managers do not possess wealth, but have some information about the statistical properties of the assets traded in the financial markets. The fiduciary managers are unable to choose asset allocations, but possess information about asset managers' estimation errors. The investors also have some idea about the asset managers' errors in general, but they are unable to distinguish between the managers with small and large errors. They can hire the fiduciary managers, who have more information about these errors, but their services are costly. The asset managers' utilities depend on the excess return over the benchmark set by the investors and the fiduciary managers' utilities depend on the total portfolio return.

The contribution of this paper is threefold. First, if fiduciary managers' information about estimation errors of the asset managers is precise, hiring the fiduciary manager is always more beneficial for the investors than delegating the portfolio management to an asset manager. Second, if the fiduciary manager can identify only the approximate scope of errors of the asset managers, then it could be optimal for the investors to delegate directly. This scenario requires that the variance of the asset manager errors is not too large. In the analysis of the equilibrium I identify the index effect, which is well documented in the literature (an increase in the price of a stock included in the index and a drop in the price of a stock excluded of the index). The second detected effect is that the prices of stocks included in the portfolio of an asset manager grow with the expansion of the fiduciary management and the prices of the stocks which are left aside decrease.

The three-tier hierarchy was introduced by Tirole, 1986 twenty five years ago. In his paper the medium tier is called a supervisor, whose role it is to collect information and to help the principal to control the agent. This is similar to the role of the fiduciary manager in my setting. Moreover, assumptions made in the current work also satisfy the axioms formulated in Tirole's paper: (1) the principal lacks either the time or the knowledge required to supervise the agent; (2) it is inefficient to have several supervisors because of the cost of duplication and collusive behaviour; (3) the supervisor lacks either the time or the resources required to run the vertical structure. However, the main focus of Tirole's paper is the possibility of a coalition between the supervisor and the agent. This behaviour is eliminated in the present study, since the model is designed in such a way that the fiduciary manager cannot directly influence the utility of the asset managers or the investor. Instead, I focus on the question of benefits and losses of the introduction of the intermediary between the principal and the agents.

Recently Portilla, 2008 has contributed to the portfolio delegation literature by extending the model of Dasgupta and Prat, 2006. In the initial model an investor hires a fund manager without knowing her trading abilities and then, observing her performance, decides whether to retain or to fire her. Portilla adds an intermediary layer of a fund company, which could be either talented (in finding out the abilities of fund managers) or untalented (which can only update her beliefs observing common information). Portilla investigates the possibilities and consequences of churning equilibria (with excessive activity of managers to generate commissions). My study extends the binary abilities of the fund manager and the fund company (good/bad, talented/untalented) to the case of continuous abilities. Moreover, these abilities might be identifiable only up to the moments of their distributions across economic agents.

Many authors consider different types of problems of decentralized investment. One of the first attempts to address the most important question in this area was made by Sharpe, 1981. He considers the situation in which more than one investment manager is employed in the process of selecting an investment portfolio and the investor can compose an optimal blend of their portfolio solutions. He distinguishes between two cases: each portfolio manager invests in a separate class of assets and all managers invest in the same group of assets. In the latter case managers' estimates of the security's expected return differ from a "consensus" estimate. Sharpe points out that in the presence of a significant disagreement about risks as well as about returns, decentralized decision making is likely to be significantly less desirable (which is coherent with my findings). The introduction of a fiduciary manager helps to improve this situation. The main difference from the present study is the key assumption of Sharpe that managers accept instructions of the investor on what function to maximize. In particular, in my setting the risk-aversion is not set by the investor and each agent maximizes her own utility functions.

One important feature of my model is the inclusion of asymmetric information and the aggregation of information in the process of its transmission between the agents. In particular, the asset managers do not share their knowledge about the asset return distribution which allows them to report only the portfolio's final return-risk profile, not the optimal portfolio weights. The fiduciary manager does not share his knowledge about errors of the asset managers with the investor, but can develop an optimal compensation scheme in exchange for a share of the investor's return. The investor does not inform the fiduciary manager that his information about asset managers is erroneous, but uses the information about the error distribution in order to choose between hiring and not hiring the fiduciary manager. The similar issue of unwillingness to share the private information is considered by Elton and Gruber, 2004. They also assume that portfolio managers will not provide a centralized manager with their return forecasts for individual securities, but will provide aggregate information about the portfolios they hold. However, to achieve an optimal overall portfolio when short sales are allowed, outside managers must be willing to supply some information and must follow instructions on how to form portfolios. Vayanos, 2003 investigates consequences of information losses due to its aggregation in a hierarchical structure, assuming that (1) agents from the lower hierarchical levels summarize their information reported to the upper level agents so that useful information is lost; (2) agents' decision problems interact. The optimal hierarchical structure depends heavily on the assumptions made, in particular, on whether there are returns to specialization.

In recent past van Binsbergen et al., 2008 consider fiduciary management problems. They assume that a fiduciary manager acts in the best interest of the beneficiaries, avoiding the three-tier incentive coordination system. They focus on the ability of the fiduciary manager to deal with the time inconsistency of the investment strategy by assuming that his utility function directly depends on the asset-to-liability ratio. Moreover, they illustrate the ability of the fiduciary manager to form a portfolio by employing multiple asset managers to invest in separate asset classes. They generalize the model assuming that the decision-maker knows only the cross-sectional distribution of the managers' risk appetites. The main conclusions drawn are: (1) a two-stage investment process can lead to additional costs in comparison to the single step investment, which is due to the difference in optimization problems, risk aversions and investment horizons; (2) an optimally designed benchmark could be the tool to mitigate these costs.

Stoughton et al., 2011 analyse a three-tier economy, focusing on rebates paid to a financial advisor by a portfolio manager. In their model the presence of the financial advisor helps small investors because otherwise they cannot afford the search costs. One of their main finding is that the kickbacks increase the use of advisory services, but negatively affect all or some investors. My setup significantly differs from theirs, since all investors in my model are identical, but their environment differs. I conclude that the financial advisor helps investors when the portfolio managers in their environment could have very erroneous portfolio allocation strategies.

The structure of the paper is as follows. Section 2 describes the model setting when fiduciary managers possess perfect information. Section 3 expands the framework and introduces fiduciary managers with imperfect information. Section 4 derives prices of assets in the equilibrium and the CAPM and characterizes the impact of the number of asset managers on the fiduciary management expansion. Section 5 presents numerical illustrations to the model and analyzes analytically intractable cases. Section 6 concludes.

2 Model Description

The economy is assumed to be populated by an infinite number of groups, uniformly distributed on the $(0, 1)$ interval. Each group consists of n asset managers, one fiduciary manager and one investor. All group members interact within their group and do not have any connection to other groups. This structure resembles a world separated into districts or cities with various qualities of financial systems, whose interaction is hampered by the geographical distance. Investors are identical across all groups, having the same initial endowment of wealth W and the risk aversion coefficient γ_I. Fiduciary managers are also identical across all groups, having the risk aversion coefficient γ_F. I will address them later as the investor and the fiduciary manager. All asset managers in the economy have different risk aversion coefficients γ_i, which remain their private information. Risk aversion coefficients are randomly drawn from a Gamma distribution with the shape parameter k and the scale parameter θ, which are known to all economic agents. The coefficients γ_i are independently distributed across all asset managers. All economic agents have the mean-variance expected utility:

$$\mathbb{E}[U(x)] = \mathbb{E}[x] - \gamma \mathbb{V}[x].$$

The investor does not have an elaborated ability to manage her wealth. The only decision she is able to make is to allocate her wealth between a risk-free bond with a return r_0 and a risky asset or a portfolio of assets, which return-risk profile is known to her. She cannot make portfolio decisions. Additionally, she knows how the compensation that she suggests to an asset manager or to the fiduciary manager affects their decisions and she is able to choose the amount of compensation to maximize her utility.

The economy has the common financial market where J assets are traded, to which all groups of agents have access. The joint multivariate distribution of their returns is characterized by a vector of mathematical expectations μ and a covariance matrix Σ. Only asset managers have information about the parameters of the multivariate distribution, but their information is not error-free, namely, the estimates of the expected returns are not precise: $\mu_i = \lambda_i^\mu \mu$. One can also assume that λ_i^μ is a matrix: this does not change the results, but complicates the notation.

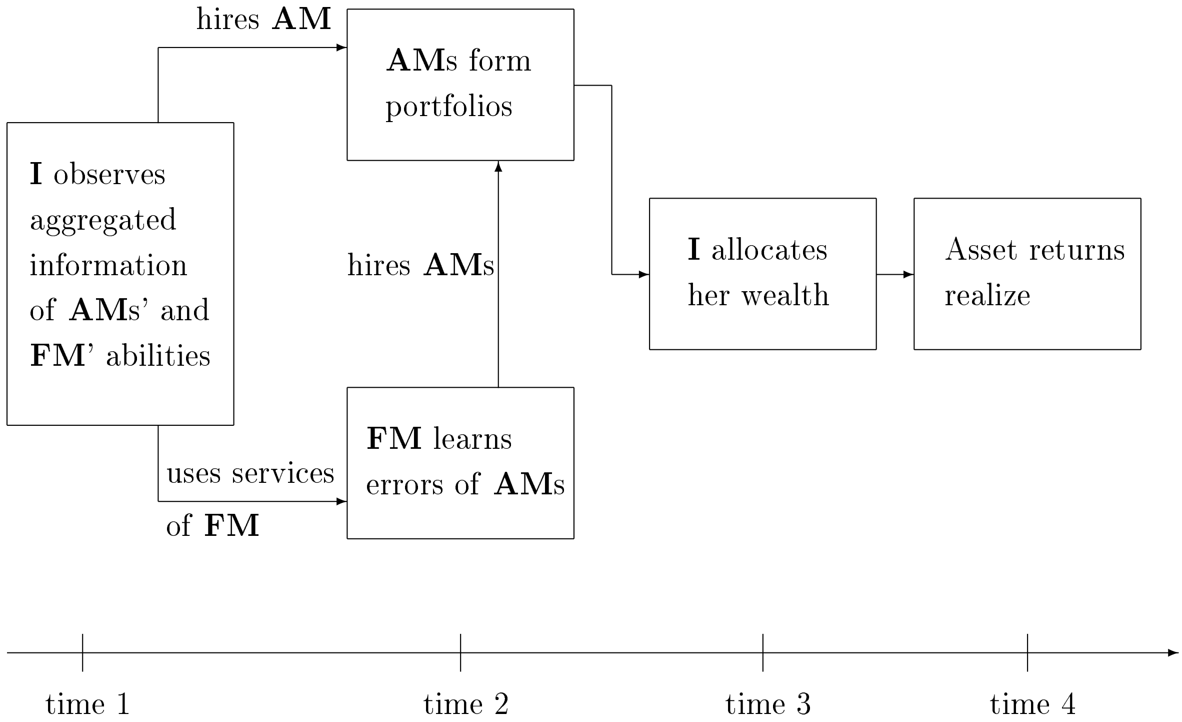
Each group $\mathcal{G}$ of agents on the unit interval is characterized by an information precision measure $g \in (0, 1)$ related to the variance of asset manager's errors λ_i^μ. The expectation of λ_i^μ is assumed to be equal to one, so that on average asset managers estimate parameters correctly, but the variance of the ratio of the estimation errors varies across the groups.

$$g = \frac{1}{\mathbb{V}[\lambda_i^\mu] + 1}, \; i \subset \mathcal{G}. \tag{1}$$

g close to one characterizes a group with asset managers having very precise estimates, whereas low g signalizes that asset managers make substantial estimation errors in their models. The value of g is the common knowledge.

All managers receive compensations of the "fulcrum" type, which means that penalties for underperforming the benchmark are symmetric to the bonuses for exceeding it. Stracca, 2006 notes in his survey that this type of contracts is typical for the mutual fund industry. This compensation structure was chosen for technical reasons, because an asymmetric call-option type compensation gives rise to a number of mathematical difficulties. For example, Cuoco and Kaniel, 2011 demonstrate how to deal with arising non-convexity and express the equilibrium as a set of non-linear equations, because the derivation of a tractable solution is impossible. Moreover, there is no agreement in the literature how to differentiate the consequences of applying a symmetric or an asymmetric type of contract. For example, Ross, 2004 shows that the hypothesis of an increase in risk-seeking due to limited liability of a manager is only partially founded, because several effects of the asymmetry may offset each other.

Figure 1: Sequence of events in the model



Note: The timeline consists of four events: (1) the investor observes some aggregate information and makes a decision about delegation; (2) a risky portfolio is suggested; (3) investment is made; (4) asset returns realize.

Figure 1 draws a timeline of events in the model. At time 1 the investor (I) observes some variables characterizing the economic environment and agents' abilities. She knows how the agents make their decisions and how they will respond to the offered compensation. Comparing values of her expected utility, the investor chooses whether to use the service of the fiduciary manager (FM) or not. At time 2, if she uses the service of the fiduciary manager, the latter offers her a contract, where it is stated how the total return will be divided between them. The fiduciary manager has bargaining power and can extract rent from the information that she has by choosing the size of the return share, which she receives for her services. If the investor decides to hire an asset manager (AM) directly, she picks a random asset manager and offers her a contract with a compensation which does not depend on the asset manager's error. If the fiduciary manager is delegated to hire asset managers, she offers all of them contracts with different compensations according to their errors. At time 3 the fiduciary manager or a chosen asset manager reports the return-risk profile of a selected portfolio to the investor and she allocates her wealth. At time 4 asset returns are realized.

2.1 The Asset Manager

A hired asset manager gets a fraction M_i of the return over a benchmark $\bar{x}$, where M_i is specified by a principal, who is either directly the investor or the delegated fiduciary manager. The investor cannot distinguish between the asset managers. Therefore, she sets $\bar{x}$ for all asset managers to be the same. Denote by R_i the return of a portfolio x_i of an asset manager i and by R_B the return of the benchmark portfolio. Then the expected utility of an asset manager i is

$$\mathbb{E}[U_i^M] = \mathbb{E}[M_i(R_i - R_B)] - \frac{1}{2}\gamma_i \mathbb{V}[M_i(R_i - R_B)]. \tag{2}$$

An asset manager maximizes the utility subject to the constraint that portfolio weights have to sum up to one (or the sum of deviations from the benchmark weights has to be zero). The overall return of the financial market is either over- or underestimated. This is captured by a factor $\lambda_i^\mu > 0$, which is the multiplier of the true asset mean vector. The degree of the deviation from the benchmark is inversely related to the compensation of an asset manager and her errors. First, the less is the compensation, the higher is the effort to outperform the benchmark. Second, the more an asset manager overestimates the expected asset returns, the more aggressively she trades with respect to the benchmark. This aggressiveness increases the expected return of the portfolio selected by an asset manager, but also raises its variance and its covariance with a portfolio of another aggressive asset manager. These statements are summarized in Proposition 2.1.

Proposition 2.1 *The optimal portfolio of an asset manager i is given by*

$$x_i = \frac{\lambda_i^\mu}{M_i\gamma_i}(\Sigma^{-1}\mu - \xi\Sigma^{-1}\boldsymbol{1}) + \bar{x}, \quad \textit{where } \xi = \frac{\boldsymbol{1}'\Sigma^{-1}\mu}{\boldsymbol{1}'\Sigma^{-1}\boldsymbol{1}}. \tag{3}$$

Corollary 2.1 *The following statistical properties hold for the return of the optimal portfolio of an asset manager i:*

- *Mathematical expectation:*
$$\mathbb{E}[R]_i = \frac{A\lambda_i^\mu}{M_i\gamma_i} + \bar{x}'\mu; \tag{4}$$
- *Covariation with the return of the optimal portfolio of an asset manager j:*

$$\mathbb{V}[R]_{ij} = \frac{1}{A}\left(\frac{A\lambda_i^\mu}{M_i\gamma_i} + B\right)\left(\frac{A\lambda_i^\mu}{M_j\gamma_j} + B\right) + \left(\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\right); \tag{5}$$

where

$$A = \mu'\Sigma^{-1}\mu - \xi\boldsymbol{1}'\Sigma^{-1}\mu, \quad B = \bar{x}'\mu - \xi.$$

The value of the expected utility at the optimum is

$$\mathbb{E}[U_i^M] = \frac{1}{2\gamma_i}A\lambda_i^\mu. \tag{6}$$

It turns out that the optimal value of the asset manager's utility does not depend on the amount of compensation. Certainly, it is due to the form of the utility function, but this is useful, since it makes the coalition of asset managers and the fiduciary manager impossible without direct transfers.

The two statistics A and B that characterize the financial market and returns of the asset managers' portfolios are assumed to be the common knowledge. Note that they are not sufficient statistics to infer the parameters of the return distribution. To find out, which signs these statistics might have, I have conducted a simulation experiment. 10000 random draws from the sample of daily returns of 325 assets with the length of 5295 days have been taken. Each draw contains 50 assets and 2000 days (time structure preserved). The benchmark portfolio is a portfolio with equal weights. All values of A and 93% of values of B are positive. Relying on this finding, I further assume that A and B are non-negative. Non-negativity here is the overall signal that asset return-risk profiles are good enough to outperform the benchmark.

2.2 The Fiduciary Manager

If the investor hires the fiduciary manager, she grants him a fraction F of the total risky return that she earns. Remember that she gets only $1 - M_i$ of the excess return earned by a portfolio of an asset manager i. Certainly, other schemes of return sharing are possible, but in this particular case the fiduciary manager, who possesses information about the asset managers' errors, can fully counterbalance them by a proper choice of the compensation M_i.

I introduce first the fiduciary manager, who has full and precise information of the errors of the asset managers in her group. She also has the ability to hire many asset managers as opposed to the investor who is able to hire only one asset manager. Tailoring compensations M_i according to λ_i^μ, she is able to compensate the impact of errors. Since asset managers are ex post identical for the fiduciary manager, it is optimal for her to invest in each portfolio $1/n$ share of the investor's wealth. Exact values of γ_i are ruled out in the expressions for the returns because the fiduciary manager takes the second mathematical expectation with respect to γ_i.

The expected utility of the fiduciary manager is equal to:

$$\mathbb{E}[U^F] = \mathbb{E}\left[\mathbb{E}\left[\frac{F}{n}\sum_{i=1}^{n} R_i(M_i)|\gamma_i\right]\right] - \frac{1}{2}\gamma_F \mathbb{E}\left[\mathbb{V}\left[\frac{F}{n}\sum_{i=1}^{n} R_i(M_i)|\gamma_i\right]\right], \tag{7}$$

where $R_i(M_i) = (1 - M_i)(R_i - R_B) + R_B$.

Proposition 2.2 *The optimal compensation M_i paid by the fiduciary manager to an asset manager i is equal to*

$$M_i^F = \left[\frac{\theta(k-1)(A - F\gamma_F B)}{AF\gamma_F \lambda_i^\mu}\frac{(k-2)n}{n(k-2)+1} + 1\right]^{-1}. \tag{8}$$

It follows from expression (8) that the more an asset manager overestimates the expected asset returns, the larger is the share of the excess return granted to him. It may seem controversial at the first glance, but this could be explained by the following argument: the variance of the excess return of an asset manager overestimating expected returns is high, therefore it is optimal for investor to have a smaller fraction of it in her portfolio in comparison to the excess return of an asset manager who estimates returns more precisely.

Taking into account Proposition 2.2 the expected utility of the fiduciary manager given the compensation scheme is:

$$\mathbb{E}[U^F] = F\bar{x}'\mu - \frac{1}{2}F^2\gamma_F\bar{x}'\Sigma\bar{x} + \frac{1}{2}\frac{\gamma_F}{A}\left(\frac{A}{\gamma_F} + FB\right)^2 \frac{n(k-2)}{n(k-2)+1}. \tag{9}$$

It is clear that the sufficient condition for $\mathbb{E}[U^F]$ to be positive (the fiduciary manager always agrees to be hired by the investor) is a moderate risk aversion of the fiduciary manager:

$$\gamma_F < \frac{2\bar{x}'\mu}{\bar{x}'\Sigma\bar{x}}. \tag{10}$$

The first two summands refer to the fraction F of the benchmark return that she receives and the last summand is obtained due to her perfect knowledge of the asset managers' errors. Maximizing this function with respect to F determines the size of the fiduciary manager's optimal compensation:

Corollary 2.2 *The optimal compensation of the fiduciary manager is*

$$F = \frac{\bar{x}'\mu - B\frac{n(k-2)}{n(k-2)+1}}{\gamma_F(\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\frac{n(k-2)}{n(k-2)+1})}. \tag{11}$$

2.3 The Investor

The investor chooses between a risk-free return r_0 and a risky portfolio with an expected return R_P and a return variance S_P^2. The general form of her expected utility function is

$$\mathbb{E}[U^I] = \beta R_P + (1-\beta)r_0 - \frac{1}{2}\gamma_I\beta^2 S_P^2. \tag{12}$$

The investor has three options to allocate her wealth. First, the investor invests directly in the benchmark. Second, the investor hires an asset manager. The investor is not able to hire more than one asset manager and is not able to distinguish between them. However, she is aware of the overall variance of the asset managers' errors $\mathbb{V}[\lambda_i^\mu]$. Therefore, the investor cannot tailor the amount of the compensation for an asset manager to her individual error. Third, the investor hires the fiduciary manager.

In order to derive a tractable solution for the size of the fiduciary manager compensation F, which is optimal for the investor, the risk-free rate is assumed to be zero.[1] Proposition 2.3 summarizes the main implications of the three options: the resulting levels of utility, the

[1] It is still possible to solve for F without this assumption, but the resulting form is cumbersome to analyse (a root of a cubic polynomial). In Section 5 we will investigate, which consequences this assumption has. The solution of the non-zero risk-free rate case is provided in Appendix B.

compensations of asset managers hired directly or through the intermediary, the optimal compensation of the fiduciary manager from the point of view of the investor and the shares of wealth invested in the risky portfolio. Clearly, the investor cannot compensate the fiduciary manager with more than the total return that she earns as well as make the fiduciary manager to pay to her for the opportunity to allocate her wealth. Moreover, it is impossible for the investor to delegate a negative share of wealth (I also assume that she cannot short the benchmark). Therefore, the proposition also states sufficient conditions to ensure $1 > F > 0$ and $\beta > 0$. While solving three different problems of the investor, it was noticed that all necessary formulae could be written in the same form if one introduces two auxiliary variables Λ and Γ and singles out an important variable κ, which measures the total agent's ability to raise the welfare of the investor.

Proposition 2.3 *The three options available to the investor can be written with the same formulas using different values of parameters κ, γ_F and λ_i:*

1. *Investment in the benchmark:* $\kappa = 0$, $\Lambda = 1$, $\Gamma = \infty$,
2. *Delegation to an asset manager:* $\kappa = \frac{k-2}{k-1}g$, $\Lambda = 1$, $\Gamma = \infty$,
3. *Delegation to the fiduciary manager:* $\kappa = \frac{n(k-2)}{n(k-2)+1}, \Lambda = \lambda_i^{\mu}, \Gamma = \gamma_F$.

The optimal value of the expected utility is

$$\mathbb{E}[U(\beta)^I] = \frac{1}{2\gamma_I}\left(\frac{(\bar{x}'\mu - B\kappa)^2}{\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}} + A\kappa\right). \tag{13}$$

The compensation of asset managers (options 2 and 3) is

$$M_i = \left[\theta(k-1)\kappa\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\Lambda(\bar{x}'\mu - B\kappa)} + 1\right]^{-1}. \tag{14}$$

The compensation of the fiduciary manager (option 3) is

$$F = \frac{\bar{x}'\mu - B\kappa}{\Gamma\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right)}. \tag{15}$$

The share of wealth invested in the risky portfolio is

$$\beta = \frac{1}{\gamma_I}\cdot\frac{1}{1-F}\cdot\frac{\bar{x}'\mu - B\kappa}{\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}}. \tag{16}$$

The expected return of the risky portfolio is

$$R_P = (1 - F)\left[\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{\bar{x}'\mu - B\kappa}\kappa + \bar{x}'\mu\right]. \tag{17}$$

The Sharpe ratio of the risky portfolio is

$$SR_P = \sqrt{\frac{(\bar{x}'\mu - B\kappa)^2}{\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}} + A\kappa} = \sqrt{2\gamma_I \mathbb{E}[U(\beta)^I]}. \tag{18}$$

The following assumptions are sufficient to have $1 > F \geq 0$ *and* $\beta \geq 0$.

1. *The market is sufficiently profitable if risk-adjusted:* $\mathbf{1}'\Sigma^{-1}\mu \geq 0$,
2. *The Sharpe ratio of the benchmark is not too high:* $\sqrt{A} > \dfrac{\bar{x}'\mu}{\sqrt{\bar{x}'\Sigma\bar{x}}}$,
3. *The fiduciary manager is sufficiently risk-averse:* $\gamma_F > \dfrac{\bar{x}'\mu}{\bar{x}'\Sigma\bar{x}}$.

The optimal value of the investor's utility function is in general the squared Sharpe ratio plus a positive constant, divided by the risk aversion of the investor. The parameter $\kappa \in [0, 1]$ summarizes the ability of all interacting agents to increase the investor's utility using their private information to create portfolio allocations, which are different from the benchmark. Since the investor does not have any information about the individual assets traded in the financial market, her κ has the lowest value of zero and her utility is proportional to the squared Sharpe ratio of the benchmark.

Since the asset managers have some imperfect information about the financial market, κ in the case of direct delegation is greater than zero. But the lower is the dispersion of the asset managers' errors, the higher is their average ability to make additional input to the investor's utility. Since the asset managers require the compensation, they decrease the nominator of the Sharpe ratio, but also decrease its denominator due to their ability to manage risks and add a positive constant benefit $A\kappa$ to the utility function.

The intermediation of the fiduciary manager pushes κ even more upwards due to her perfect information about asset managers' errors and the ability to delegate portfolio management to many asset managers at the same time that creates benefits of diversification. These two effects could be separated by comparing the structure of κ_2, which is provided by the asset managers only, and κ_3, which is provided by the joint work of the fiduciary manager and asset managers:

$$\kappa_2 = \kappa_3 \left(1 - \frac{n-1}{n(k-1)}\right) g = \kappa_3 \times \text{Loss of no diversification} \times \text{Loss of error variance}. \quad (19)$$

It could be noticed that the loss of the asset managers' error variance amplifies the loss of no diversification. If there is a fiduciary manager, then the more asset managers are here, the higher is the benefit of diversification. In this case κ is closer to one but the marginal benefit of diversification is decreasing. However, it is not very plausible that the fiduciary manager can learn about the errors of all asset managers if their number is high. We will address this issue in the next section.

It is important that the fiduciary manager is not too less risk-averse, because in this case the reward-risk profile of the suggested portfolio could be inappropriate to the investor. The benchmark provided by the investor should not have a too high Sharpe ratio, because then asset managers should take too risky positions trying to outperform it and will spoil the qualities of their portfolios. These two facts are reflected in the sufficient assumptions to have reasonable values of the fiduciary manager compensation and the share invested in the risky portfolio.

We observe also that it is not beneficial for the fiduciary manager to use values of λ_i^μ that are different from the true ones, because she receives a part of the return retained by the investor. The higher is the quality of the return, the more she retains. The maximum in quality is reached when the fiduciary manager fully eliminates the impact of the asset managers' errors choosing an appropriate compensation based on the true λ_i^μ. Therefore, she cannot obtain additional returns to pay some transfers to asset managers and coalitions are impossible.

It is very important that the optimal value of the compensation chosen by the fiduciary manager coincides with the size of the compensation, which the investor would choose herself. It means that the contract suggested by the fiduciary manager is optimal for the investor. This convenient result is due to the assumption of the zero risk-free rate. In Section 3 we discuss in detail the situation when the optimal values of F for the investor and the fiduciary manager are different.

2.4 Comparative Statics

In this subsection we briefly consider how the variables characterizing the economy respond to changes in exogenous parameters and compare the investor's expected utility functions of the three investment options.

Corollary 2.3 *The sensitivity of the variables listed in Proposition 2.3 with respect to the parameter κ is as follows:*

- *Utility function of the investor:*

$$\frac{\partial \mathbb{E}[U(\beta)^I]}{\partial \kappa} = \frac{A}{2\gamma_I}\left(\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{A\bar{x}'\Sigma\bar{x} - B^2\kappa}\right)^2 \geq 0.$$

- *Compensation of an asset manager:*

$$\frac{\partial M_i}{\partial \kappa} = -\left[\theta(k-1)\kappa\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\Lambda(\bar{x}'\mu - B\kappa)} + 1\right]^{-2}\theta(k-1)\bar{x}'\mu\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{A\Lambda(\bar{x}'\mu - B\kappa)^2} < 0.$$

- *Compensation of the fiduciary manager:*

$$\frac{\partial F}{\partial \kappa} = -\frac{BA}{\Gamma}\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{[A\bar{x}'\Sigma\bar{x} - B^2\kappa]^2} \leq 0.$$

- *Share of wealth invested in the risky portfolio:*

$$\frac{\partial \beta}{\partial \kappa} = -\frac{B}{\gamma_I A}\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{[\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A} - \frac{1}{\gamma_F}(\bar{x}'\mu - B\kappa)]^2} \leq 0.$$

- *Total return from the risky part of the investment:*

$$\frac{\partial \beta R_P}{\partial \kappa} = \frac{A}{\gamma_I}\left(\frac{A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu}{A\bar{x}'\Sigma\bar{x} - B^2\kappa}\right)^2 \geq 0$$

- *Sharpe ratio of the risky portfolio:*

$$\frac{\partial SR_P}{\partial \kappa} = \frac{1}{2}\sqrt{\frac{A[A\bar{x}'\Sigma\bar{x} - B\bar{x}'\mu]^4}{(A\kappa\bar{x}'\Sigma\bar{x} - 2B\kappa\bar{x}'\mu + (\bar{x}'\mu)^2)(A\bar{x}'\Sigma\bar{x} - B^2\kappa)^3}} \geq 0$$

The first important fact is that if the fiduciary manager has perfect information about the asset managers' errors, then her presence increases the total ability of agents to increase the investor's utility. In this case the investor always chooses Option 3. In Section 3 we consider the fiduciary manager with imperfect information and Option 3 will not always be preferred.

Second, the asset managers' compensation decreases as their total investment ability increases. The investor gives asset managers a smaller proportion of her return in order to push the risky portfolio further away from the benchmark and benefit from the improvement in the information quality. Asset managers are not aware of their errors, therefore, the compensation is not a carrot but a stick. Since the compensation of the fiduciary manager and the compensation which she pays asset managers are positively related (see Proposition 2.2), the share of the return paid to the fiduciary manager also decreases with the increase in the information quality. The increased return share that the investor retains allows her to invest less in the risky asset to maintain her desired risk-return profile. The increase in the total investment ability of the agents also increases the Sharpe ratio of the risky portfolio.

3 Imperfect Information of the Fiduciary Manager

In this section we consider the fiduciary manager with a reduced ability to account for the asset managers' errors. In this case, the fiduciary manager fails to obtain the full information about λ_i^μ. The erroneous estimates $(\lambda_i^\mu)^\varepsilon$ of the fiduciary manager are

$$(\lambda_i^\mu)^\varepsilon = \frac{\lambda_i^\mu}{\varepsilon_i}, \quad \mathbb{E}[\varepsilon] = \mathbf{1}, \, \mathbb{V}[\varepsilon] = \Xi, \, \mathbb{P}[\varepsilon_i < 0] = 0. \tag{20}$$

The fiduciary manager realizes that her information is not perfect, but is not able or willing to improve her estimates of the errors. The investor knows the mean and the covariance matrix of the fiduciary manager's errors. These errors change the statistical characteristics of the portfolio that the fiduciary manager offers, which has a substantial impact on the investor's expected utility.

We assume that the errors of the fiduciary manager do not depend on the individual characteristics of the asset managers. Therefore, the variance of ε_i is constant across all i. Similarly, the correlations between ε_i and ε_j are set to be constant, i.e. $\frac{\Xi_{ij}}{\Xi_{ii}} = \rho$. To preserve positive definiteness of the matrix Ξ, ρ must be greater than $-\frac{1}{n-1}$. It is reasonable to assume that, as the number of asset managers grows, it becomes harder to gather the important information to estimate their errors, therefore, the variance of ε_i is an increasing function of n denoted by $\Xi_{ii} = \zeta = \zeta(n)$.

The optimization problem of the fiduciary manager remains the same as in equation (7) and, therefore, her desirable F also does not change. On the contrary, the investor takes into account the errors in the fiduciary manager's information, which is reflected in her optimal F. Hence, the optimal F of the investor and the fiduciary manager are not identical now.

Proposition 3.1 *The optimal compensation from the viewpoint of the fiduciary manager is*

$$F^F = \frac{\bar{x}'\mu - B\kappa}{\gamma_F\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right)}, \tag{21}$$

and the optimal compensation from the viewpoint of the investor is

$$F^I = \frac{\bar{x}'\mu - B\kappa\left(1 - \frac{\bar{x}'\mu}{B}\chi\right)}{\gamma_F\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\left(1 - \frac{\bar{x}'\mu}{B}\chi\right)\right)}, \tag{22}$$

where

$$\kappa = \frac{n(k-2)}{n(k-2)+1}, \quad \chi = \frac{\zeta[(k-1)+(n-1)(k-2)\rho]}{n(k-2)}.$$

Proposition 3.1 shows that the investor would like to force the fiduciary manager to be more risk averse (F^I is increasing in χ), since the quality of the offered portfolio is now worse. However, if the investor still finds it optimal to use the services of the fiduciary manager, the fiduciary manager knows that her information is valuable and she requires her optimal compensation F^F. If Option 3 dominates Option 2 in a group g even with this size of the compensation, the investor accepts the contract. If Option 2 becomes more attractive for the investor, she begins to bargain and pushes F closer to her optimal F^I. The bargaining process ends either if F_I^ε is reached or if the investor is indifferent between Option 2 and Option 3.[2]

Corollary 3.1 *If the fiduciary manager is erroneous, then for Option 3 the following holds.*

The value of the expected utility is

$$\mathbb{E}[U(\beta_3)^I](F) = \frac{1}{2\gamma_I}\frac{\left[A\kappa\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right) + \bar{x}'\mu\right]^2}{A\kappa\left(\frac{1}{F^2\gamma_F^2} - \frac{B^2}{A^2}\right) + A\kappa^2\chi\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right)^2 + \bar{x}'\Sigma\bar{x}}. \tag{23}$$

The share of wealth invested in a risky portfolio is

$$\beta_3(F) = \frac{1}{\gamma_I(1-F)}\frac{A\kappa\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right) + \bar{x}'\mu}{A\kappa\left(\frac{1}{F^2\gamma_F^2} - \frac{B^2}{A^2}\right) + A\kappa^2\chi\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right)^2 + \bar{x}'\Sigma\bar{x}}. \tag{24}$$

[2] The investor chooses Option 3 in this case, because the fiduciary manager can always set F by a tiny amount closer to F^I, which makes Option 3 strictly better for the investor.

The compensation of the fiduciary manager is

$$F = \begin{cases} F^F = \dfrac{\bar{x}'\mu - B\kappa}{\gamma_F\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right)} & \text{if } \mathbb{E}[U(\beta_2)^I] \leq \mathbb{E}[U(\beta_3)^I](F^F), \\ F : \mathbb{E}[U(\beta_2)^I] = \mathbb{E}[U(\beta_3)^I](F) & \text{if } \mathbb{E}[U(\beta_3)^I](F^F) \leq \mathbb{E}[U(\beta_2)^I] \leq \mathbb{E}[U(\beta_3)^I](F^I)), \\ 0 & \text{if } \mathbb{E}[U(\beta_2)^I] > \mathbb{E}[U(\beta_3)^I](F^I). \end{cases} \tag{25}$$

given that $F^I < 1$, which is equivalent to $\chi < \left[1 + \frac{\bar{x}'\Sigma\bar{x}\gamma_F - \bar{x}'\mu}{B\kappa(1-\gamma_F\frac{B}{A})}\right]\frac{B}{\bar{x}'\mu}$.

Clearly, if the variance of the fiduciary manager's errors is small, there are still gains from hiring the fiduciary managers for all groups. If the variance is high, this does not add anything valuable to the investor's information, but distorts it. However, for the variance values within a certain interval there is a separation: in some groups of agents with a higher variance of the asset managers' errors investors prefer to hire the intermediary, and in other groups they do not. Conditions for such a separation to exist and the exact formula for the index of the separating group are given in Proposition 3.2.

Proposition 3.2 *If $\underline{\zeta} < \zeta(n) < \overline{\zeta}$, where*

$$\underline{\zeta} = \{\zeta : \max_g \mathbb{E}[U(\beta_2)^I] = \mathbb{E}[U(\beta)^I](F^F)\},$$

$$\overline{\zeta} = \{\zeta : \min_g \mathbb{E}[U(\beta_2)^I] = \mathbb{E}[U(\beta)^I](F^I)\},$$

then there exists a $g^* \in (0, 1)$,

$$g^* = \frac{\kappa}{1 + \kappa\chi}\frac{k-1}{k-2}, \tag{26}$$

where

$$\kappa = \frac{n(k-2)}{n(k-2)+1}, \quad \chi = \frac{\zeta[(k-1)+(n-1)(k-2)\rho]}{n(k-2)},$$

such that the investor in groups with $g \leq g^*$ prefers to hire the fiduciary manager and the investor in groups with $g > g^*$ directly hires a random asset manager.

4 General Equilibrium

In order to derive the equilibrium of the model, I assume the per capita net[3] supply of assets S_j, $j = 1, \ldots, J$ to be exogenous and fixed.[4] As derived in the previous section, there exist certain conditions, under which some investors hire the fiduciary manager and some investors directly hire a random asset manager. Denote by P_j the price of the jth asset and by Q_{ij}^g the demand for an asset j of a manager i in a group g. The share x_{ij}^g, which the asset manager i allocates to an asset j in the group g is rewritten as:

$$x_{ij}^g = \frac{P_j Q_{ij}^g}{\tilde{\beta} W}$$

$\tilde{\beta}$ for Option 2 is equal to the share β that the investor allocates to the risky portfolio, and it is equal to β/n if the investor chooses Option 3. Denote by $i^\dagger$ the index of a randomly chosen asset manager.

The total demand for an asset j must be equal to the supply of an asset j. One should distinguish between the demand in groups with g smaller than the separation level g^*, where the fiduciary manager's services are needed, and in the rest of the economy, where the investors channel their money directly to the asset managers. To find the total demand, one integrates with respect to g.

$$\int_0^{g^*} \sum_{i=1}^{n} Q_{ij}^g \, dg + \int_{g^*}^{1} Q_{i^\dagger j}^g \, dg + \bar{Q}_j = S_j \tag{27}$$

Note that this equation can be applied in the systems, where the investors in all groups choose Option 3: it corresponds to $g^* = 1$. The price that balances demand and supply can be easily found.

Proposition 4.1 *The equilibrium price for the asset j is*

$$P_j = \frac{W}{S_j - \bar{Q}_j} \left[\left(\bar{x}_j + \frac{\pi_j}{(k-1)\theta} \right) \breve{\boldsymbol{\beta}} + \pi_j C \right], \tag{28}$$

[3]Due to the construction of the model the total demand for the asset i could be negative, which would lead to a negative equilibrium price. To avoid this, I assume that there is a high enough positive demand for each asset $\bar{Q}_j$. For example, this could be the demand of noise traders. If so, the sign of the net supply coincide with the sign of the total demand of the asset managers.

[4]A similar approach is used in Leippold and Rohner, 2011, with the difference that in their paper the price does not reveal the full information because of stochastic component of the supply, and in the current paper the asset demand is stochastic.

$$\pi = \Sigma^{-1}\mu - \xi\Sigma^{-1}\boldsymbol{1},$$

$$\breve{\boldsymbol{\beta}} = \beta_3(F^F)\hat{g} + \int_{F^F}^{F^I} \beta_3(F)\,dF + \int_{g^*}^{1} \beta_2(g)\,dg,$$

$$C = \kappa_3\hat{g}\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa_3} + \kappa_3\int_{F^F}^{F^I}\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right)\beta_3(F)\,dF + \int_{g^*}^{1}\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa_2(g)}\kappa_2(g)\,dg,$$

$$\hat{g} = \{g : \mathbb{E}[U(\beta_2(g))^I] = \mathbb{E}[U(\beta)^I](F_F^{\varepsilon})\}.$$

It follows that the price before the normalization by $S_j - \bar{Q}_j$ is equal to the sum of two terms. The first term is the average wealth that would be invested in an asset j by asset managers if they retained the whole return of the portfolios delegated to them. The second term is the "correction" term, which arises from the attempts of the fiduciary managers and the investors to influence the choice of portfolio managers.

If the benchmark is a market index, then the weights $\bar{x}_j$ are usually positive. Then the model encompasses the same index effect as demonstrated in Basak and Pavlova, 2011. If a stock l is added to the index ($\Delta\bar{x}_l > 0$) and a stock k is dropped ($\Delta\bar{x}_k < 0$), then the price of the stock l gets a boost and that of the stock k falls, given the overall level of index does not change after the replacement. We can see from the expression in (28) that the sign of the price effect coincides with the sign of $\Delta\bar{x}_j$. This relation was also obtained by Cuoco and Kaniel, 2010.

Under an additional assumption it is possible to derive the CAPM.

Proposition 4.2 *If* $\frac{S_j}{S_j - \bar{Q}_j}$ *are equal across all* j*, then the following CAPM equation holds for the economy:*

$$\mathbb{E}[R] - r_0\mathbf{1} = \beta^{CAPM}(\mathbb{E}[R^M] - r_0), \tag{29}$$

$$\beta^{CAPM} = \frac{\left[\frac{1}{(k-1)\theta} + \frac{C}{\breve{\boldsymbol{\beta}}}\right](\mu - \xi\boldsymbol{1}) + \Sigma\bar{x}}{A\left[\frac{1}{(k-1)\theta} + \frac{C}{\breve{\boldsymbol{\beta}}}\right]^2 + 2B\left[\frac{1}{(k-1)\theta} + \frac{C}{\breve{\boldsymbol{\beta}}}\right] + \bar{x}'\Sigma\bar{x}}. \tag{30}$$

The government can influence the expansion of the fiduciary management through the parameter n, by hampering or stimulating the activity of the asset managers. For example, it is possible to shrink the expansion of the fiduciary management by regulating the number of asset managers. If the elasticity of the fiduciary managers' error variance is low, this number must be decreased. The exact condition on the elasticity is stated below.

Proposition 4.3

$$\frac{\partial \zeta}{\partial n}\frac{n}{\zeta(n)}\left(1+\frac{n\rho(k-2)}{(k-1)-(k-2)\rho}\right) < 1+\frac{1}{\zeta[(k-1)-(k-2)\rho]} \quad \Leftrightarrow \quad \frac{\partial g^*}{\partial n} > 0. \tag{31}$$

5 Numerical Analysis

To illustrate the main findings of the model, the model is calibrated to real data. The data consist of daily observations on 325 stocks, which are components of S & P 500 index or were these during some period of time. I take the investment horizon of one year and I set the portfolio, in which all stocks have equal weights, as the benchmark. The estimates of the mean vector and the covariance matrix of returns are simply their unconditional empirical counterparts. It results in the following return-risk profile of the benchmark: the yearly expected return $\bar{x}'\mu$ is 17.45%, the yearly variance $\bar{x}'\Sigma\bar{x}$ is 0.0337, the Sharpe ratio of the benchmark is 0.9514. Two sufficient statistics of the economy are: $A = 5.6934$ and $B = 0.0608$. For these parameter values the first two assumptions of Proposition 2.3 are satisfied and the risk aversion of the fiduciary manager must belong to the interval (5.19, 10.38).

For the basic setting I assume that the number of asset managers in each group is 10, the parameters of the Gamma distribution are: $\theta = 1$ and $k = 5$, which result in the average degree of the risk aversion of asset managers equal to 5. The risk aversion of the investor γ_I is set to 20 and the risk aversion of the fiduciary manager γ_F is set to 10. These values of the risk aversion are consistent with their empirical estimates (see e.g. Haug et al., 2011, Hagiwara and Herce, 1997). I assume that the investor is the most risk-averse agent, since she knows almost nothing about the financial market and cares about her wealth, therefore, she puts a significant weight on the return variance. The asset managers believe that they know a lot, therefore they may be more aggressive and less risk-averse. Finally, I make the fiduciary manager moderately risk-averse.

First, I illustrate the implications of the model for different values of k, g and n in the case when the fiduciary manager has precise information. I consider the group with $g = 0.5$ as the basic case. The results of the simulation are presented in Table 1 and Table 2.

Table 1: Direct hiring of an asset manager

	$g = 0.1$	$g = 0.5$	$g = 0.9$	$k = 3$	$k = 7$	$k = 10$
Expected utility	0.0321	0.0706	0.1095	0.0545	0.0760	0.0795
Size of risky allocation	25.25%	22.68%	20.07%	23.75%	22.32%	22.08%
Asset manager compensation	94.68%	76.06%	60.83%	90.92%	65.21%	53.67%
Net portfolio return	25.45%	62.25%	109.09%	45.89%	68.07%	72.06%
Portfolio Sharpe ratio	1.1337	1.6802	2.0926	1.4766	1.7430	1.7837

Note: The table presents simulated results of direct hiring of a random asset manager in three different groups and for three different shape parameters of the risk-averseness distribution. The first row shows the expected utility of the investor, the second row shows the proportion of the investor's wealth invested in the risky portfolio. The third row corresponds to the share of the excess return given to the hired asset manager. The last two rows present the net return and the Sharpe ratio of the risky portfolio chosen by the asset manager.

Table 2: Using services of the fiduciary manager

	$n = 2$	$n = 10$	$n = 25$	$k = 3$	$k = 7$	$k = 10$
Expected utility	0.1333	0.1479	0.1504	0.1401	0.1495	0.1505
Size of risky allocation	29.30%	26.92%	26.53%	28.17%	26.66%	26.51%
Fid. manager compensation	36.95%	35.00%	34.66%	36.04%	34.78%	34.65%
Net portfolio return	90.99%	109.85%	113.39%	99.51%	112.19%	113.55%
Portfolio Sharpe ratio	2.3092	2.4321	2.4527	2.3677	2.4457	2.4536

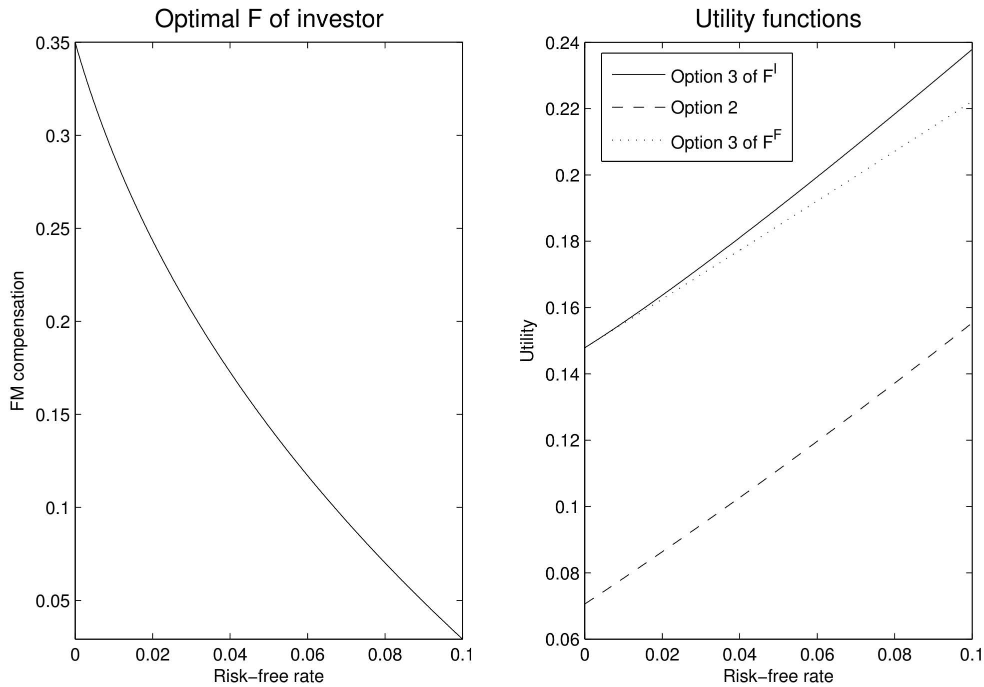
Note: The table presents simulated results of using the service of the fiduciary manager for three different numbers of asset managers in a group and for three different shape parameters of the risk-averseness distribution. The first row shows the expected utility of the investor, the second row shows the proportion of the investor's wealth invested in the risky portfolio. The third row corresponds to the share of the total return given to the fiduciary manager. The last two rows present the net return and the Sharpe ratio of the risky portfolio presented by the fiduciary manager.

The values of expected returns are high, because asset managers do not face any constraints on portfolio weights, they can have unlimited leverage and are not imposed to any transaction costs. However, the Sharpe ratios of the resulting portfolios are not enormously high. We can notice that if the fiduciary manager has precise information, a change in n or k has only a tiny impact on the optimal allocation. In the case of direct hiring of asset managers the size of the risky allocation also remains very stable, the most responding variable is the compensation of asset managers. In the basic setting the switch from Option 2 to Option 3 leads to the 18%

increase of the size of the risky allocation and the Sharpe ratio rises by about 45%.

Next, I consider the consequences of the zero risk-free rate assumption. The details of the solution in this case are described in Appendix B, but it is impossible to analyse the reaction of the optimal compensation of the fiduciary manager and the expected utility analytically. Therefore, I make a graphical illustration of how different values of the risk-free rate affect these quantities. The maximum possible yearly risk-free rate is assumed to be 10%.

Figure 2: Impact of the risk-free rate



Note: The figure illustrates how the fiduciary manager compensation, optimal for the investor, and the investor's utility under different circumstances depend on the risk-free rate. The dashed line corresponds to the expected utility from Option 2, the dotted line to the expected utility from Option 3 and the solid lines to the expected utility from Option 3 if the investor set F. The information of the fiduciary manager is precise.

Figure 2 shows that a non-zero risk-free rate lowers the expected utility of Option 3. Clearly, the utility of F optimal for the investor is higher that the utility of F set by the fiduciary manager (it corresponds to F of the zero risk-free rate case). But we see that despite the increasing gap between the preferences of the investor and the fiduciary manager, the utility of the investor from hiring the fiduciary manager is still higher than that from the direct delegation.

Additionally, the 1% increase of the risk-free rate lowers F optimal for the investor by 0.35%. The explanation for this observation is that if the risk-free rate grows, the investor cares less about the quality of the risky portfolio and allows asset managers to deviate more from the benchmark. This increases the risky return, but lowers the Sharpe ratio. The risk-free rate also increases the expected utility.

To consider the case of the erroneous fiduciary manager, I set the basic value of ρ to 0.1. I also assume that ζ has the following functional form: $\zeta(n) = 0.1\sqrt{n^3}$. I illustrate the consequences of the fiduciary manager's errors by considering different values of n and ρ. I do not show results for different k, since it has only a tiny impact.

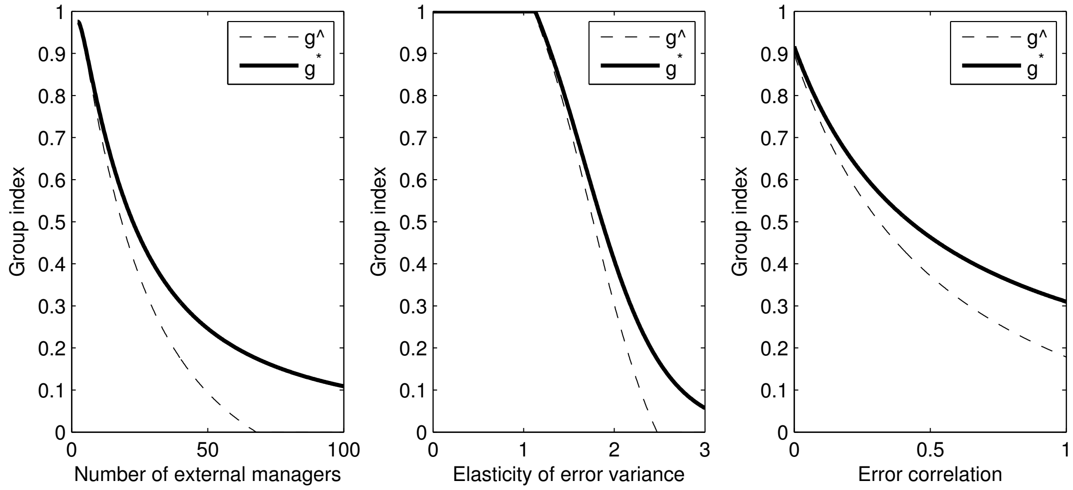
We see from Table 3 that in the basic case after adding errors to the estimates of the fiduciary manager the expected utility has dropped. The size of the risky allocation has decreased by 37% in the basic case and the Sharpe ratio by 21%. The quality of the portfolio deteriorates remarkably fast as the number of asset managers or the error correlations grow.

Table 3: The case of the erroneous fiduciary manager

	$n=2$	$n=10$	$n=25$	$\rho=0$	$\rho=0.4$	$\rho=0.7$
Expected utility	0.1163	0.0930	0.0706	0.1094	0.0706	0.0706
Size of risky allocation	25.55%	16.93%	22.68%	19.91%	66.32%	22.68%
Fiduciary manager compensation	36.95%	35.00%	0.00%	35.00%	74.98%	0.00%
Net portfolio return	90.99%	109.85%	62.25%	109.84%	21.27%	62.25%
Portfolio Sharpe ratio	2.1565	1.9287	1.6802	2.0914	1.6802	1.6802

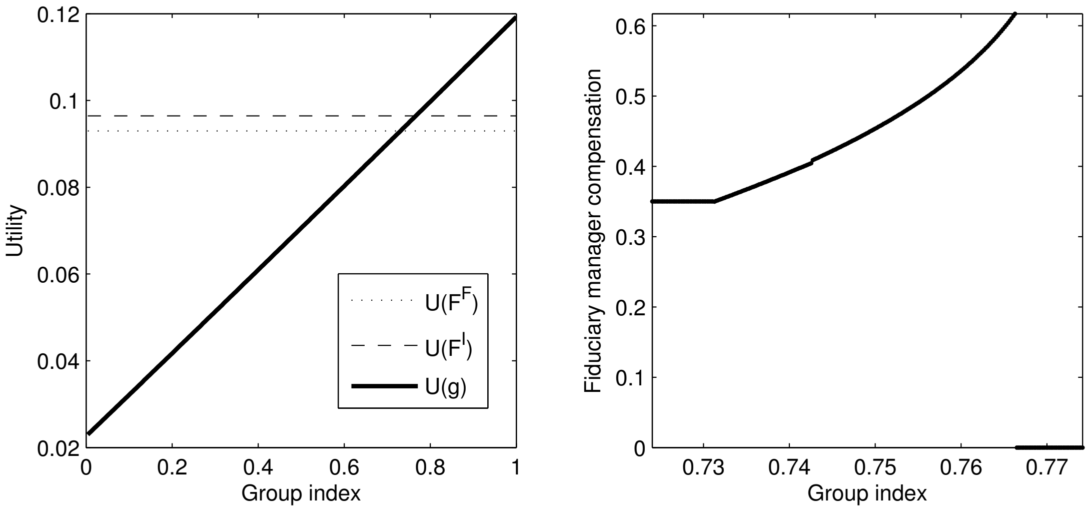
Note: The table presents simulated results of using the service of the erroneous fiduciary manager for three different numbers of asset managers in a group and for three different values of the error correlations. The first row shows the expected utility of the investor, the second row shows the proportion of the investor's wealth invested in the risky portfolio. The third row corresponds to the share of the total return given to the fiduciary manager. The last two rows present the net return and the Sharpe ratio of the risky portfolio presented by the fiduciary manager.

Figure 3: Impact of the number of asset managers, the error variance elasticity and the error correlation on the interval of bargaining



Note: The figure shows how the interval of bargaining limited by the group indexes g^* and $\hat{g}$ changes with the number of asset managers in a group n, the error variance elasticity $\frac{\partial \zeta}{\partial n} \frac{n}{\zeta(n)}$ and the error correlation ρ. The dashed lines correspond to $\hat{g}$, the solid lines to g^*.

Figure 4: Impact of the quality of asset managers in a group on the investor's expected utility and the fiduciary manager's compensation



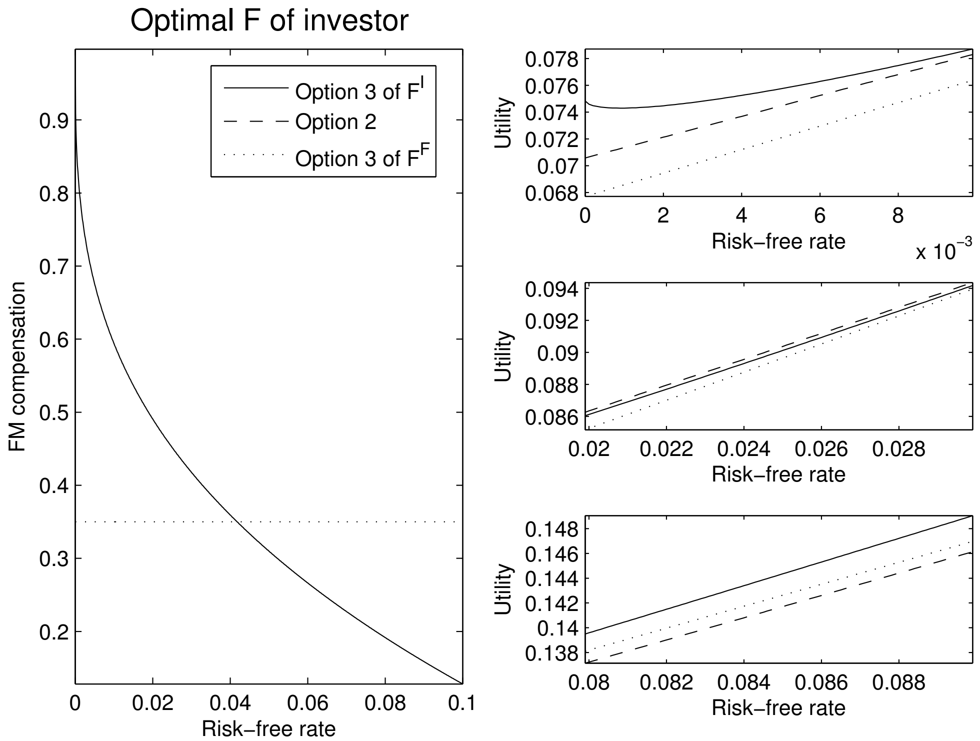
Note: The left part of the figure shows how the value of the expected utility of the investor depends on the group of asset managers she is dealing with. The dashed line corresponds to $[U(\beta_3)^I](F^I)$, the dotted line to $[U(\beta_3)^I](F^F)$ and the solid line to $[U(\beta_2)^I](g)$. The right part of the figure shows how the value of the fiduciary manager compensation changes as the quality of the asset managers in a group rises.

Figure 3 shows the impact of the number of asset managers, the elasticity of the error variance with respect to n and the error correlations on g^* and $\hat{g}$. The area below the dashed line corresponds to the values of g, for which the fiduciary manager is hired and her desired value of F is accepted. The area between the dashed line and the solid line is the values of g, for which the investor uses services of the fiduciary manager, but bargains about F. For g's above the solid line the investor prefers to delegate directly. Clearly, as the values of considered variables increase, the quality of the fiduciary manager's services decays and less groups prefer her services. The bargaining area also extends.

The left part of Figure 4 shows the relation between utilities in different groups. The final graph of the utility would be the maximum of the dotted and solid lines. The right part of Figure 4 shows the dynamics of F in the bargaining area. If g increases by 0.01 in this area, the fiduciary manager receives about 7% more of the total return.

The last case from Section 3 to consider is the case of the erroneous fiduciary manager and a non-zero risk-free rate. To demonstrate the most interesting case, I set ρ to 0.35. The left part of Figure 5 shows that for lower risk-free rates F, which is optimal for the investor, is greater than F, which is optimal for the fiduciary manager, and for higher risk-free rates the opposite holds. The relation between utilities leads to three different cases. In the first case, when $r_0 \in [0, 0.0144]$ and $r_0 \in [0.0362, 0.0370]$ (the upper right graph of Figure 5), Option 2 is preferred to Option 3 with $F = F^F$, but Option 3 with $F = F^I$ is preferred to Option 2. Therefore, the area between the dashed and the dotted lines is the bargaining area. For $r_0 \in [0.0144, 0.0362]$ Option 2 dominates Option 3. For $r_0 \in [0.0370, 0.1]$ Option 3 with $F = F^F$ dominates Option 2 and the investor accepts F set by the fiduciary manager. The reason for this is that the additional return from the risk-free rate makes the investor more tolerable to the errors of the fiduciary manager.

Figure 5: Impact of the risk-free rate on the fiduciary manager compensation and the investor's expected utility

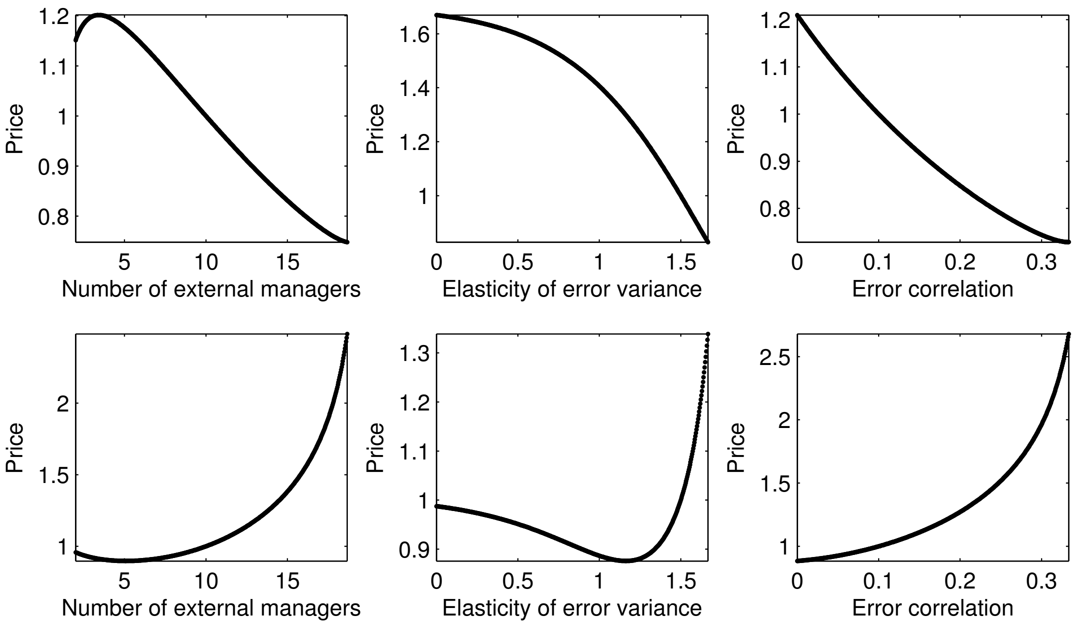


Note: The left part of the figure shows how the optimal compensation of the fiduciary manager set by the investor changes with the risk-free rate. The information of the fiduciary manager in this case contains errors. The right part presents three possible orders of the utility curves (depending on particular values of the risk-free rate) for Option 2 and Option 3 with Fs optimal for the investor / the fiduciary manager. The dashed lines correspond to Option 2, the dotted lines to Option 3, the solid lines to Option 3 if the investor set F.

Finally, I investigate the behaviour of prices in the general equilibrium, since it is not possible to pursue analytically. Computing of $\breve{\boldsymbol{\beta}}$ and C for the basic case shows that C is much greater than $\breve{\boldsymbol{\beta}}$ (0.2916 and 3.7540 respectively). Therefore, the effect of C mostly dominates in the dynamics of stock prices. The only case, when the effect of C is eliminated, is when the weight of a stock in the portfolio is equal to zero. I consider the cases of $\pi_j = 0$ and $\pi_j = 0.1$. The price is normalized so that it is equal to one in the basic case.

Figure 6 displays the dynamics of the price when either the number of asset managers or the elasticity of error variances or error correlations change. We see that most of the time the price of the stock with the zero weight in the portfolio moves in the opposite direction in comparison to the price of the other stock. It means that if the fiduciary management expands due to the improvement of the quality of its services or due to a decrease of the number of asset managers, the prices of the stocks included in the asset manager's portfolio grow and the prices of other stocks sink. Basak and Pavlova, 2011 also find that stock prices grow in the presence of the institutional investors but their investors hold all stocks in their portfolios.

Figure 6: Impact of three different factors on the price of a stock



Note: The figure shows how the stock prices change with the number of asset managers in a group n, the error variance elasticity $\frac{\partial \zeta}{\partial n} \frac{n}{\zeta(n)}$ and the error correlation ρ. The three upper figures correspond to the case of $\pi_j = 0.1$ (the stock is included in the benchmark), the three lower figures to the case of $\pi_j = 0$ (the stock is not in the benchmark).

6 Conclusion

I analyse the economy with the three-tier hierarchy of economic agents: investors could delegate the control of their wealth to fiduciary managers, who in turn hire asset managers to make portfolio decisions. The novel part of the paper is to study the impact of the amount and precision of information, which varies across agents, on the investor's decision about portfolio delegation. All three types of agents maximize their own benefits, which is sporadic in the literature on the portfolio delegation. Moreover, the bargaining behaviour of the investor and the intermediary is analyzed in situations, in which the first party can threaten to switch to the direct delegation and the second party requires a compensation for valuable information.

In the benchmark case, which I consider, the fiduciary manager possesses perfect information about drawbacks in the asset managers' strategies. In this situation he can design contracts that fully compensate the informational imperfections of the asset managers. Obviously, the investors always prefer to hire fiduciary managers. After that I extend the setup and consider the fiduciary manager with imperfect information. While in the benchmark case an increase in the number of asset managers is profitable for the fiduciary manager, because the gains from diversification grow, the imperfect information environment reduces the precision of his information and works against him by decreasing the quality of his services. In groups with badly informed asset managers the filtering done by fiduciary managers still brings some additional gains to investors. Therefore, their services are demanded. But in groups, where the information of the asset managers has only minor errors, such filtering adds more noise than extracts valuable information, inducing investors to channel their money directly to asset managers.

Considering the general equilibrium of the model I detected two effects. First, the well-known index effect: if the benchmark is the index, then the price of a stock added to the index rises and the price of an excluded stock falls. Second, the prices of assets included in portfolios of asset managers grow with the expansion of fiduciary management and the prices of assets which are left aside sink. The government can influence the expansion of the fiduciary management by hampering or stimulating the activity of asset managers. The policy depends on the elasticity of the error variance of the fiduciary manager.

Further research on this topic can consider measures, which the fiduciary manager would take to improve his performance in the case of imperfect information. It would be reasonable for him to decrease the number of asset managers that he hires. It leads to a competition between asset managers and possible money transfers to the fiduciary manager. But due to the estimation errors of the asset managers, the one who agrees to pay the highest price to be hired ex-ante is not the best one ex-post. It would be interesting to investigate the impact of this phenomenon on the decision of the investors.

7 Appendix

Appendix A. Proofs

Proof of Proposition 2.1

An asset manager i assumes that asset returns have a multivariate normal distribution with a mean vector $\lambda_i^\mu \mu$ and a covariance matrix Σ. Therefore, the portfolio with asset weights x has a normal distribution with the mean $\lambda_i^\mu x'\mu$ and the variance $x'\Sigma x$ (see Theorem 2.16, p.43 Fang et al., 1990). If $\bar{x}$ are the weights of a benchmark and Δx are deviations from the benchmark, then an asset manager i solves the following optimization problem:

$$\max_{\Delta x,\, \Delta x'\mathbf{1}=0} M_i \lambda_i^{\mu} \Delta x' \mu - \frac{1}{2} M_i^2 \gamma_i \Delta x' \Sigma \Delta x, \; \Delta x = x_i - \bar{x}. \tag{32}$$

Differentiation of the corresponding Lagrangian gives the following first order conditions.

$$\lambda_i^{\mu} \mu - M_i \gamma_i \Sigma \Delta x - \eta \mathbf{1} = 0, \; \Delta x' \mathbf{1} = 0. \tag{33}$$

Solving for Δx and expressing x_i as the sum of Δx and $\bar{x}$ leads to equation (3).

Proof of Proposition 2.2

Denote by R the vector of portfolio returns of asset managers and by γ the vector of the risk aversion coefficients. Straightforward computation gives the expressions for its first raw and second centered moments:

$$\mathbb{E}[\mathbb{E}[R|\gamma]]_i = \frac{A\lambda_i^{\mu}(1-M_i)}{(k-1)\theta M_i} + \bar{x}'\mu; \tag{34}$$

$$\mathbb{E}[\mathbb{V}[R|\gamma]]_{ij} = \begin{cases} \frac{A(\lambda_i^{\mu})^2(1-M_i)^2}{(k-1)(k-2)\theta^2 M_i^2} + \frac{2B\lambda_i^{\mu}(1-M_i)}{(k-1)\theta M_i} + \bar{x}'\Sigma\bar{x} & \text{if } i = j, \\ \frac{A\lambda_i^{\mu}\lambda_j^{\mu}(1-M_i)(1-M_j)}{(k-1)^2\theta^2 M_i M_j} + \frac{B(1-M_i)\lambda_i^{\mu}}{(k-1)\theta M_i} + \frac{B(1-M_j)\lambda_j^{\mu}}{(k-1)\theta M_j} + \bar{x}'\Sigma\bar{x} & \text{if } i \neq j. \end{cases} \tag{35}$$

Substitution of (34) and (35) in (7), differentiation with respect to each M_i and rearranging give the following set of the first order conditions.

$$\sum_j \frac{\lambda_j^{\mu}(1-M_j)}{(k-1-\mathbf{1}_{j=i})M_j} = \frac{n\theta(A - F\gamma_F B)}{AF\gamma_F} \quad \forall i \tag{36}$$

It is easy to show that the solutions of this set of equations are given by equation (8).

Proof of Proposition 2.3

Maximizing (12) with respect to β we get an expression of the optimal share of wealth invested in the risky portfolio given its expected return and return variance:

$$\beta = \frac{R_P - r_0}{\gamma_I S_P^2}. \tag{37}$$

The expected utility of this β is equal to

$$\mathbb{E}[U(\beta)^I] = r_0 + \frac{1}{2}\frac{(R_P - r_0)^2}{\gamma_I S_P^2}. \tag{38}$$

The expected return and the return variance of the benchmark portfolio are $\bar{x}'\mu$ and $\bar{x}'\Sigma\bar{x}$ respectively. Substituting them to (37) and (38) and setting r_0 to zero give expressions (16) and (13) if $\kappa = 0$ and $\Gamma = \infty$. Obviously, expressions (17) and (18) also hold true.

Option 2 implies channeling money to one of asset managers without the diversification opportunity. The expected return and the return variance of this option are

$$R_P = \frac{A(1-M)}{M(k-1)\theta} + \bar{x}'\mu, \quad S_P^2 = \frac{A(1-M)^2}{M^2 g(k-1)(k-2)\theta^2} + \frac{2B(1-M)}{M(k-1)\theta} + \bar{x}'\Sigma\bar{x}. \tag{39}$$

Substituting these expressions in (38) and maximizing with respect to M, we can solve for the optimal M (the first order condition turns out to be a linear equation) given in (14) for $\kappa = \frac{k-2}{k-1}g$ and $\Lambda = 1$. Equations (13) and (16)-(18) follow after substitution of the optimal M and rearrangement of terms.

For the last option the fiduciary manager offers the investor the portfolio with the following profile:

$$R_P = (1-F)\left(\left(\frac{A}{F\gamma_F} - B\right)\frac{n(k-2)}{n(k-2)+1} + \bar{x}'\mu\right); \tag{40}$$

$$S_P^2 = (1-F)^2\left(\left(\frac{A}{F^2\gamma_F^2} - \frac{B^2}{A}\right)\frac{n(k-2)}{n(k-2)+1} + \bar{x}'\Sigma\bar{x}\right). \tag{41}$$

Repeating the steps made to solve for Option 2, but maximizing the expected utility with respect to F, we derive equation (15). Equations (13) and (16)-(18) follow after substitution of the optimal F and rearrangement of terms.

To prove the sufficiency of the three listed assumptions, first notice that from (15) and (16) $1 > F \geq 0$ implies $\beta \geq 0$. From the positive definiteness of the covariance matrix Σ and the first assumption it follows that $\xi \geq 0$. Therefore, we can write the nominator of (15) as

$$\bar{x}'\mu - B\kappa = \bar{x}'\mu - (\bar{x}'\mu - \xi)\kappa = \bar{x}'\mu(1-\kappa) + \xi\kappa. \tag{42}$$

It was assumed that $B \geq 0$. Therefore, $\bar{x}'\mu \geq 0$ and $\bar{x}'\mu - B\kappa \geq 0$, since it is a convex combination of two nonnegative terms. Therefore, for F to be nonnegative the denominator of

(15) must be positive. It could be rewritten as

$$\Gamma\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right) = \Gamma\left(\bar{x}'\Sigma\bar{x} - \frac{(\bar{x}'\mu)^2\kappa}{A} + \frac{\xi(\bar{x}'\mu + B)}{A}\right). \tag{43}$$

The expression reaches its minimum for $\kappa = 1$, since if $\bar{x}'\Sigma\bar{x} - \frac{(\bar{x}'\mu)^2}{A}$ is positive, the denominator is positive. Rearranging $\bar{x}'\Sigma\bar{x} - \frac{(\bar{x}'\mu)^2}{A} > 0$ leads to the second assumption.

$1 > F$ is equivalent to $\Gamma > \frac{\bar{x}'\mu - B\kappa}{\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}}$. The derivative of the right-hand side is equal to $-B\frac{\bar{x}'\Sigma\bar{x} - \frac{(\bar{x}'\mu)^2\kappa}{A} + \frac{\xi\bar{x}'\mu}{A}}{\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right)^2}$, which is negative because of the first and second assumptions. Therefore, the right-hand side reaches its maximum for $\kappa = 0$ and the third assumption is sufficient for $\Gamma > \frac{\bar{x}'\mu - B\kappa}{\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}}$ to hold for all κ.

Proof of Proposition 3.1

For the last option the fiduciary manager offers the investor the portfolio with the following profile:

$$R_P = (1 - F)\left(\left(\frac{A}{F\gamma_F} - B\right)\kappa + \bar{x}'\mu\right); \tag{44}$$

$$S_P^2 = (1 - F)^2\left(\left(\frac{A}{F^2\gamma_F^2} - \frac{B^2}{A}\right)\kappa + A\chi\left(\frac{\kappa}{F\gamma_F} - \frac{B}{A}\kappa\right)^2 + \bar{x}'\Sigma\bar{x}\right). \tag{45}$$

$$\kappa = \frac{n(k-2)}{n(k-2)+1}, \quad \chi = \frac{\zeta[(k-1) + (n-1)(k-2)\rho]}{n(k-2)}. \tag{46}$$

These values could be computed from (34) and (35) using $\frac{\lambda_i^\mu}{\varepsilon_i}$ and taking one more mathematical expectation with respect to ε. χ comes from the additional error terms ε, which could not be compensated by the fiduciary manager, its general formula is

$$\chi = \frac{(k-2)\mathbf{1}'\Xi\mathbf{1} + \mathrm{tr}[\Xi]}{n^2(k-2)}. \tag{47}$$

Substituting these variables in (38) and maximizing with respect to F lead to equation (22).

Proof of Proposition 3.2

$\underline{\zeta} < \zeta(n) < \overline{\zeta}$ is sufficient to ensure that $\exists g : \mathbb{E}[U(\beta_2(g))^I] < \mathbb{E}[U(\beta_3)^I](F^F)$, so that Option

3 is preferred and $\exists g : \mathbb{E}[U(\beta_2(g))^I] > \mathbb{E}[U(\beta_3)^I](F^I)$, so that Option 2 is preferred.

g^* is the solution of $\mathbb{E}[U(\beta_2(g))^I] = \mathbb{E}[U(\beta_3)^I](F^I)$ since there is no bargaining opportunities for the fiduciary manager to make Option 3 more attractive after this point. Substituting (22) in (23) and simplifying, we get the following expression for $\mathbb{E}[U(\beta_3)^I](F^I)$ (κ and χ as in (47)):

$$\mathbb{E}[U(\beta_3)^I](F^I) = \frac{1}{2\gamma_I}\left(\frac{(\bar{x}'\mu - B\frac{\kappa}{1+\kappa\chi})^2}{\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\frac{\kappa}{1+\kappa\chi}} + A\frac{\kappa}{1+\kappa\chi}\right). \tag{48}$$

The expression has the same form as (13), therefore $g^* = \frac{\kappa}{1+\kappa\chi}\frac{k-1}{k-2}$.

Proof of Proposition 4.1

The general form of the demand for an asset j of an asset manager i in a group g is

$$Q_{ij}^g = \left[\frac{\lambda_i^\mu}{M_i\gamma_i}\pi_j + \bar{x}_j\right]\tilde{\beta}W, \tag{49}$$

where $\pi = \Sigma^{-1}\mu - \xi\Sigma^{-1}\mathbf{1}$.

Three different cases for β and M are possible: (1) $g < \hat{g}, \hat{g} :\ \mathbb{E}[U(\beta_2(\hat{g}))^I] = \mathbb{E}[U(\beta)^I](F^F)\}$, when investors accept services of the fiduciary manager and her price F^F; (2) $\hat{g} \le g \le g^*$, when investors accept services of the fiduciary manager, but negotiate about F, F depends on g of a group; (3) $g > g^*$, when investors delegate directly, M depends on g of a group.

Substituting for Q_{ij}^g, accounting for differences across g and appropriate rearrangement, we get the following equation for the price P_j:

$$P_j = \frac{W}{S_j - \bar{Q}_j}\Upsilon_j, \tag{50}$$

where

$$\Upsilon_j = \bar{x}_j\mathbb{E}\left[\beta_3(F^F)\hat{g} + \int_{\hat{g}}^{g^*}\beta_3(F(g))\,dg + \int_{g^*}^{1}\beta_2(g)\,dg\right]$$

$$+\pi_j\mathbb{E}\left[\beta_3(F^F)\hat{g}\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda_i^\mu}{M_i(F^F)\gamma_i} + \int_{\hat{g}}^{g^*}\frac{1}{n}\sum_{i=1}^{n}\frac{\beta_3(F(g))\lambda_i^\mu}{M_i(F(g))\gamma_i}\,dg + \frac{\lambda_{i^\dagger}^\mu}{\gamma_{i^\dagger}}\int_{g^*}^{1}\frac{\beta_2(g)}{M_2(g)}\,dg\right].$$

We can notice the following facts: (1) since the mapping from g to F is bijective on $[\hat{g}, g^*]$, integrating on this interval by g is equivalent to integrating on the interval $[F^F, F^I]$ by F; (2)

since γ_i, ε_i and λ_i^μ are independent of all other random variables and of each other, we can replace them in the formula by their expectations doing sequential conditioning. It leads to the final expression for Υ_j:

$$\Upsilon_j = \left(\bar{x}_j + \frac{\pi_j}{(k-1)\theta}\right)\left[\beta_3(F^F)\hat{g} + \int_{F^F}^{F^I} \beta_3(F)\,dF + \int_{g^*}^{1} \beta_2(g)\,dg\right]$$

$$+\pi_j\left[\kappa_3\hat{g}\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa_3} + \kappa_3\int_{F^F}^{F^I}\left(\frac{1}{F\gamma_F} - \frac{B}{A}\right)\beta_3(F)\,dF + \int_{g^*}^{1}\frac{\bar{x}'\Sigma\bar{x} - \frac{B}{A}\bar{x}'\mu}{\bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa_2(g)}\kappa_2(g)\,dg\right].$$

Proof of Proposition 4.2

The weight of an asset j in the market portfolio is equal to

$$x_j^M = \frac{P_jS_j}{\sum_{j=1}^{J} P_jS_j} = \bar{x}_j + \pi_j\left[\frac{1}{(k-1)\theta} + \frac{C}{\breve{\boldsymbol{\beta}}}\right] = \bar{x}_j + \pi_j\boldsymbol{C}. \tag{51}$$

The well-known formula for the beta of CAPM is $\beta_j = \dfrac{\mathbb{C}[r_j, R^M]}{\mathbb{V}[R^M]}$, where r_j is the return of the asset j and R^M is the return of the market portfolio.

Straightforward computation gives

$$\mathbb{V}[R^M] = \bar{x}'\Sigma\bar{x} + 2\pi'\Sigma\bar{x}\boldsymbol{C} + \pi'\Sigma\pi\boldsymbol{C}^2 = \bar{x}'\Sigma\bar{x} + 2B\boldsymbol{C} + A\boldsymbol{C}^2 \tag{52}$$

$$\mathbb{C}[R^M, r] = \Sigma\bar{x} + \Sigma\pi\boldsymbol{C} = \Sigma\bar{x} + (\mu - \xi\mathbf{1})\boldsymbol{C} \tag{53}$$

Proof of Proposition 4.3

Differentiation of equation (26) with respect to n shows that the sign of $\frac{\partial g^*}{\partial n}$ is defined by the sign of $\frac{\partial \kappa}{\partial n} - \frac{\partial \chi}{\partial n}\kappa^2$. This expression is equal to

$$\frac{k-2}{(n(k-2)+1)^2}\left[1 - \zeta'(n)n[(k-1) + (n-1)(k-2)\rho] + \zeta(n)[(k-1) - (k-2)\rho]\right] \tag{54}$$

$$\frac{\partial \zeta}{\partial n}\frac{n}{\zeta(n)}\left(1 + \frac{n\rho(k-2)}{(k-1) - (k-2)\rho}\right) > \frac{1}{\zeta[(k-1) - (k-2)\rho]} + 1 \quad \Rightarrow \quad \frac{\partial g^*}{\partial n} < 0 \tag{55}$$

Appendix B. Non-zero risk-free rate

Equations (13)-(18) still hold for options 1 and 2 if one replaces $\bar{x}'\mu$ by $\bar{x}'\mu - r_0$. It is due to the fact that the investor retains the return from the benchmark and simply considers the excess return over the risk-free rate to make her decisions. For Option 3 it is not so, since $(1-F)$ of $\bar{x}'\mu$ also goes to the fiduciary manager.

In this case substitution of (40) and (41) in (38) and differentiation with respect to F do not lead to a simple linear equation, it has an additional nonlinear summand multiplied by the risk-free rate r_0:

$$\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A} - \frac{\bar{x}'\mu - B\kappa}{\gamma_F F} + r_0\left[\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}\right)\frac{\gamma_F}{A\kappa}\frac{F^2}{(1-F)^2} + \frac{1}{\gamma_F F(1-F)^2}\right] = 0 \tag{56}$$

Making the following renotation

$$K_1 = \bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa, \quad K_2 = \bar{x}'\mu - B\kappa \tag{57}$$

and simplifying equation (56), one gets the following cubic equation for F:

$$K_1\gamma_F\left(\frac{\gamma_F}{A\kappa}r_0 + 1\right)F^3 - (2K_1\gamma_F + K_2)F^2 + (K_1\gamma_F + 2K_2)F + r_0 - K_2 = 0. \tag{58}$$

This equation has an analytical solution, the exact formula could be found e.g. in Irving (2004), p.149.

In this case the optimal F's for the investor and for the fiduciary manager do not coincide, since the investor has an additional source of return: the risk-free rate, and is less sensitive to the portfolio of the fiduciary manager. Therefore, she would be willing to give him a smaller F in order to push him slightly further from the benchmark than in the case of zero risk-free rate.

In the case of the erroneous fiduciary manager equation (56) has an additional summand:

$$\begin{aligned} &\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa Z(\chi)}{A} - \frac{\bar{x}'\mu - B\kappa Z(\chi)}{\gamma_F F} + \\ &r_0\left[\left(\bar{x}'\Sigma\bar{x} - \frac{B^2\kappa}{A}(1-\chi\kappa)\right)\frac{\gamma_F}{A\kappa}\frac{F^2}{(1-F)^2} + \frac{1+\chi\kappa}{\gamma_F F(1-F)^2} - \frac{B\chi\kappa(1+F)}{A(1-F)^2}\right] = 0, \end{aligned} \tag{59}$$

where $Z(\chi) = 1 - \frac{\bar{x}'\mu}{B}\chi$.

Making the following renotation

$$K_1(\chi) = \bar{x}'\Sigma\bar{x} - \frac{B^2}{A}\kappa Z(\chi), \quad K_2(\chi) = \bar{x}'\mu - B\kappa Z(\chi) \quad K_3 = \gamma_F \chi \kappa \frac{B}{A} \tag{60}$$

and simplifying equation (59), one gets the following cubic equation for F:

$$\left(K_1(\chi)\gamma_F + \frac{r_0}{\chi\kappa}K_3^2 + r_0\frac{\gamma_F^2}{A\kappa}K_1(0) \right) F^3 - (2K_1(\chi)\gamma_F + K_2(\chi) + r_0 K_3)F^2 + \tag{61}$$
$$(K_1(\chi)\gamma_F + 2K_2(\chi) - r_0K_3)F + r_0(1+\chi\kappa) - K_2(\chi) = 0.$$

References

[1] S. Basak and A. Pavlova. Asset prices and institutional investors. February 2011.

[2] BlackRock. A short introduction to fiduciary management. Professional Pensions, February 2010.

[3] D. Cuoco and R. Kaniel. Equilibrium prices in the presence of delegated portfolio management. *Journal of Financial Economics*, 101:264–296, 2011.

[4] A. Dasgupta and A Prat. Financial equilibrium with career concerns. *Journal of Theoretical Economics*, 88:288–307, 2006.

[5] E. Elton and M. Gruber. Optimum centralized portfolio construction with decentralized portfolio management. *Journal of Financial and Quantitative Analysis*, 39(3):481–494, 2004.

[6] K.-T. Fang, S. Kotz, and Ng K.W. *Symmetric multivariate and related distributions.* Chapman and Hall (London and New York), 1990.

[7] M. Hagiwara and M. Herce. Risk aversion and stock price sensitivity to dividends. *American Economic Review*, 87(4):738–745, 1997.

[8] J. Haug, T. Hens, and P. Wöhrmann. Risk aversion in the small and in the large. June 2011.

[9] R. S. Irving. *Integers, polynomials, and rings: a course in algebra.* Springer-Verlag New York, 2004.

[10] M. Leippold and P. Rohner. Equilibrium implications of delegated asset management under benchmarking. *The Review of Finance*, forthcoming, 2011.

[11] Goldman Sachs Asset Management. Is fiduciary management the answer? Sponsored Comment, 2005.

[12] Y. Portilla. Two-sided career concern and financial equilibrium. December 2008.

[13] S. Ross. Conmpensation, incentives, and the duality of risk aversion and riskiness. *Journal of Finance*, 59(1):207–225, 2004.

[14] W.F. Sharpe. Decentralized investment management. *Journal of Finance*, 36:217–234, 1981.

[15] N. Stoughton, Y. Wu, and J. Zechner. Intermediated investment management. *Journal of Finance*, 66(3):947–980, 2011.

[16] L. Stracca. Delegated portfolio management: A survey of the theoretical literature. *Journal of Economic Surveys*, 20(4):823–848, 2006.

[17] J. Tirole. Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, and Organization*, 2(2):181–214, 1986.

[18] J.H. van Binsbergen, M. W. Brandt, and R. S. J. Koijen. Optimal decentralized investment management. *Journal of Finance*, 63(4):1849–1895, 2008.

[19] D. Vayanos. The decentralization of information processing in the presence of interactions. *Review of Economic Studies*, 70:667–695, 2003.

Mixture Normal Conditional Correlation Models

This paper was presented at

- FINRISK Research Day 2010, Gerzensee, Switzerland (June 2010)
- 4th International Conference on Computational and Financial Econometrics, London, UK (December 2010)
- Statistisch-Ökonometrisches Seminar, Christian-Albrechts-Universität zu Kiel, Kiel, Germany (May 2012)
- 66th European Meeting of the Econometric Society, Malaga, Spain (August 2012)

Abstract

I propose a class of hybrid models to describe and predict the dynamics of a multivariate stationary random vector, e.g. a vector of stock returns. These models combine essential features of the multivariate mixture normal distribution and the conditional correlation models. I describe in detail the expectation-maximization algorithm, which makes the parameter estimation feasible and fast virtually for any random vector length. I fit the suggested models to five data sets, consisting of vectors of stock returns, with the maximum vector length of fifteen stocks. The predictive ability of this model class is compared to that of other widely used multivariate models, and it turns out that my models provide the best forecasts, both on average and for extreme negative returns. All necessary formulas to apply these models for important financial objectives are also provided.

Keywords: Finite Mixtures; Dynamic Conditional Correlation; Forecasting; Predictive Ability; Multivariate Modelling

JEL classification: C51; C53; G17

1 Introduction

Modeling and predicting the dynamics of a multidimensional set of asset returns is a very important prerequisite for many financial decisions and actions, such as construction and evaluation of an optimal portfolio, pricing of structured products that are designed with more than one underlying asset or calculation of a hedge ratio. Certainly, the most desirable object is a predicted multivariate density of the return vector, since all the popular risk measures, such as variance, Value-at-Risk and Expected Shortfall, could be calculated on its basis and useful interval forecasts could be done.

The popular in modeling of financial time series ARCH and GARCH frameworks have already been extended to the multivariate case: the VEC model of Bollerslev et al., 1988 and the BEKK model of Engle and Kroner, 1995. However, an application of this type of models in their general formulation is possible only for small-dimensional vectors or under additional restrictions, e.g, the widely used diagonal BEKK. To avoid the curse of dimensionality, the family of conditional correlation models has been developed. The underlying idea is to exploit the decomposition of the covariance matrix into the "sandwich" product of the diagonal matrices of standard deviations and the correlation matrix. Bollerslev, 1990 introduces the constant conditional correlation model (CCC), which assumes that the correlation matrix is constant over time. Tse and Tsui, 2002, Engle, 2002 and Engle and Kelly, 2012 propose some dynamic conditional correlation (DCC) models and specify possible processes for the conditional correlation matrix. Despite some drawbacks and critique, this family has wide empirical application. For further details on multivariate GARCH models see the survey of Bauwens et al., 2006.

Recalling the fact that the removal of the time-varying covariance component from the residuals does not result in Gaussian residuals (Bollerslev et al, 1992), we can expect potential losses in the quality of the forecast density when applying models with the assumption of normality. Pesaran and Pesaran, 2007 suggest to use a t distribution instead of a Gaussian one, and their data favour t-DCC specification. Galeano and Ausin, 2010 use a mixture of normal distributions to fit devolatized returns. Haas et al., 2004 and Alexander and Lazar, 2006 introduce a univariate finite mixture distribution with time-varying second moments. Each component has its own dynamics of this moment. In a few years the univariate framework has been generalized to the multivariate one by Bauwens et al, 2007 and Haas et al., 2009. Both papers state that the VEC and BEKK specifications are serious restrictions for this model to be applied to the case when the number of assets is higher than, say, three.

The main contribution of this article is resolving the curse of dimensionality for finite multivariate mixture models. This is done by combining the salient features of the conditional correlation models and finite mixture distributions to formulate a class of mixture normal conditional correlation models. Kon, 1984 argues that a finite mixture of Gaussian distributions could be a good approximation of the true distribution of the asset returns. He notices that returns behave differently during information arrivals and in non-informative periods and on different weekdays. Another example is that the stock returns depend on both firm-specific and market wide information. There are also advantages to use mixture models for its mathematical flexibilities: it can produce multi-modal and fat-tailed densities, and due to its additive structure it is not difficult to investigate its properties. To remove the limit on the number of jointly modeled assets, Bauwens et al, 2007 and Haas et al., 2009 propose to employ the conditional correlation family. Additional argument for this family is the finding of the difference in correlation coefficients across regimes in Haas et al., 2009.

Using the statistical tests for comparing density forecasts suggested by Giacomini and White, 2006 and Diks et al., 2011, I show that the suggested class of models significantly outperforms a number of competing models, for which parameter estimation is feasible in the case of a large number of assets. The parameters of mixture normal conditional correlation models could be estimated using a numerically reliable and fast multi-step EM algorithm, which is described in details. The MCMC algorithm could also be applied here, but is difficult to implement for a large number of assets due to a heavy computational burden. Finally, I derive closed-form expressions for many formulas, which are required to apply the model class for a wide variety of financial goals, such as risk management, portfolio optimization and investment performance evaluation.

The remainder of this paper is as follows. Section 2 describes in detail the construction of the mixture normal conditional correlation models, Section 3 explains the algorithm of its estimation. Section 4 shows applications of the model in finance. Section 5 analyses the data and Section 6 evaluates the performance of the model using five data sets with different number of assets and discusses the results. Section 7 concludes.

2 Model Family

Let $\boldsymbol{r}_t = (r_{1t}, \ldots, r_{nt})'$ be a n-dimensional vector of returns, calculated as the first difference of logarithms of asset prices p_{it}: $r_{it} = \log p_{it} - \log p_{i,t-1}$. The vector of returns is decomposed into the sum of a conditional mean vector and an innovation vector:

$$\boldsymbol{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t. \tag{1}$$

$\boldsymbol{\mu}_t$ is the vector of conditional means given the information set $\Omega_{t-1} = \{\boldsymbol{r}_{t-1}, \boldsymbol{r}_{t-2}, \dots\}$ available at time t, so that $\mathbb{E}[\boldsymbol{\varepsilon}_t | \Omega_{t-1}] = 0$.

Since there is a strong evidence of little or no dependence of the conditional mean on the conditioning information set (the efficient market hypothesis, see Fama, 1970 and Fama, 1991), I do not address the question of theoretical modelling of the conditional mean. In the closely related papers of Bauwens et al., 2007 and Haas et al., 2009 the conditional mean vector is assumed to be constant. Other authors do AR(1) prefiltering of returns (Jacquier et al., 2002) or even allow AR(d) dynamics (Bollerslev et al., 2009).

I assume that $\boldsymbol{\varepsilon}_t$ follows a conditional multivariate finite mixture normal distribution with G components. Each component possesses a mean vector, which is assumed to be constant, and a dynamic covariance matrix. I consider three ways to model the dynamics of the covariance matrices: the CCC model (Bollerslev, 1990), the DCC model (Engle, 2002 and Engle, 2009) and the DECO model (Engle and Kelly, 2012). It is possible to use any other model listed in Bauwens et al, 2006, but I focus on these three models. These models have a common methodology of separating the covariance matrix into the product of standard deviations and correlations and modelling these two objects separately. The CCC model assumes a constant conditional correlation matrix, the DCC model assumes that the correlation matrix follows an ARMA-type process. The DECO model suggests to average the correlations of matrices in the DCC process to get the so-called equicorrelation matrix, in which all off-diagonal elements are the same.

Among these three models, only the estimates of parameters of the CCC model are proven to be consistent and asymptotically normal (see Francq and Zakoïan, 2010 for the detailed proof). There is a working paper of Engle and Sheppard, 2001, where they try to prove the same properties for the DCC parameter estimates, but fail to do so because of ignoring the fact that the target correlation matrix is also estimated. Engle, 2009 writes up the missing moment condition (11.27) on p. 135, which, however, does not always hold if we look at the unconditional expectation of the last equation of (11.24) on p.134. Aielli, 2011 shows that the DCC estimator of the target correlation matrix can be inconsistent and suggests a correction for the DCC model. He also provides a heuristic proof of consistency of the estimator of the corrected DCC model. For my model class I adopt the corrected version of the DCC model suggested by Aielli, 2011. Engle and Kelly, 2012 also use this version and Engle, 2009 mentions it as a possible solution.

The conditional density of $\boldsymbol{\varepsilon}_t|\Omega_{t-1}$ could be written as

$$f(\boldsymbol{x};\boldsymbol{\Psi}) = \sum_{g=1}^{G} \lambda_g \phi(\boldsymbol{x};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_{g,t}), \quad \lambda_g > 0 \ \forall g, \quad \sum_{g=1}^{G} \lambda_g = 1, \quad \sum_{g=1}^{G} \lambda_g \boldsymbol{\mu}_g = \mathbf{0}. \tag{2}$$

where $\phi(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the density of the multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{x}$. $\boldsymbol{\Psi}$ combines all the parameters of the finite mixture distribution, namely, $(\lambda_1,\ldots,\lambda_{G-1})$, $(\boldsymbol{\mu}_1',\ldots,\boldsymbol{\mu}_{G-1}')$ and the parameters of the processes for the covariance matrices, which will be specified later. λ_G is defined from the constraint on the component probabilities: $\sum_{g=1}^{G}\lambda_g = 1$; and $\boldsymbol{\mu}_G$ is defined from the constraint on the zero conditional mean of $\boldsymbol{\varepsilon}_t$: $\sum_{g=1}^{G}\lambda_g\boldsymbol{\mu}_g = \mathbf{0}$.

The covariance matrix of a component g is represented as the following product of three matrices:

$$\boldsymbol{\Sigma}_{g,t} = \boldsymbol{D}_{g,t}\boldsymbol{C}_{g,t}\boldsymbol{D}_{g,t}. \tag{3}$$

The matrix $\boldsymbol{D}_{g,t}$ is a diagonal matrix containing conditional standard deviations of ε_{it} in each state of the world g. Their dynamics are described with the following GARCH(p,q) process:

$$\boldsymbol{D}^2_{g,ii,t} = \omega_{g,i} + \sum_{j=1}^{p}\alpha_{g,i,j}\varepsilon^2_{i,t-j} + \sum_{k=1}^{q}\beta_{g,i,k}\boldsymbol{D}^2_{g,ii,t-k}. \tag{4}$$

Note that the sets of the GARCH parameters ($\omega_{g,i},\ \alpha_{g,i,j},\ \beta_{g,i,k},\ j=1,\ldots,p,\ k=1,\ldots,q$) are different across the assets and the components. The suggested GARCH-type model is the most common one, any other univariate conditional heteroscedasticity model can be applied to model $\boldsymbol{D}^2_{g,ii,t}$.

Three different models for the dynamics of the correlation matrix of a component g are:

- Constant conditional correlations:

$$\begin{aligned} \boldsymbol{\eta}_{g,t} &= \boldsymbol{D}^{-1}_{g,t}(\boldsymbol{\varepsilon}_{g,t} - \boldsymbol{\mu}_g), \\ \boldsymbol{C}^{CCC}_{g,t} &= \mathbb{E}[\boldsymbol{\eta}_{g,t}\boldsymbol{\eta}'_{g,t}]. \end{aligned} \tag{5}$$

- Dynamic conditional correlations:

$$\begin{aligned}
\boldsymbol{\eta}_{g,t} &= \boldsymbol{D}_{g,t}^{-1}(\boldsymbol{\varepsilon}_{g,t} - \boldsymbol{\mu}_g), \\
\tilde{\boldsymbol{\eta}}_{g,t} &= (\boldsymbol{Q}_{g,t} \odot \boldsymbol{I}_n)^{\frac{1}{2}} \boldsymbol{\eta}_{g,t}, \qquad (6) \\
\boldsymbol{Q}_{g,t} &= (1 - \sum_{j=1}^{r} \gamma_g - \sum_{k=1}^{s} \delta_g)\bar{\boldsymbol{Q}}_g + \sum_{j=1}^{r} \gamma_{g,j} \tilde{\boldsymbol{\eta}}_{g,t-j} \tilde{\boldsymbol{\eta}}'_{g,t-j} + \sum_{k=1}^{s} \delta_{g,k} \boldsymbol{Q}_{g,t-k}, \qquad (7) \\
\bar{\boldsymbol{Q}}_g &= \mathbb{E}[\tilde{\boldsymbol{\eta}}_{g,t} \tilde{\boldsymbol{\eta}}'_{g,t}], \quad \bar{\boldsymbol{Q}}_g \odot \boldsymbol{I}_n = \boldsymbol{I}_n, \qquad (8) \\
\boldsymbol{C}_{g,t}^{DCC} &= (\boldsymbol{Q}_{g,t} \odot \boldsymbol{I}_n)_{g,t}^{-\frac{1}{2}} \boldsymbol{Q}_{g,t} (\boldsymbol{Q}_{g,t} \odot \boldsymbol{I}_n)_{g,t}^{-\frac{1}{2}}. \qquad (9)
\end{aligned}$$

- Dynamic equicorrelations:

$$\begin{aligned}
\rho_{g,t} &= \frac{1}{n(n-1)} (\mathbf{1}'_n \boldsymbol{C}_{g,t}^{DCC} \mathbf{1}_n - n), \qquad (10) \\
\boldsymbol{C}_{g,t}^{DECO} &= \rho_{g,t} \boldsymbol{I}_n + (1 - \rho_{g,t}) \mathbf{1}_n \mathbf{1}'_n. \qquad (11)
\end{aligned}$$

$\odot$ is the component-wise product of two matrices, $\mathbf{1}_n$ is the column vector of ones of length n and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

The key idea is to model the individual dynamics of the variances and the correlation structure for each component g of the mixture. It would, for example, correspond to a different nature of dependence between assets in the periods of boom and recession. This extension makes the model very flexible, even in the case of only two components.

3 Parameter Estimation

General guidance on the estimation of different finite mixture models is given in the book of McLachlan and Peel, 2000. There is a variety of estimation approaches such as graphical methods, method of moments, minimum-distance methods, maximum likelihood and Bayesian approaches. The authors of the book emphasize fitting mixture models by maximum likelihood estimation and the expectation-maximization (EM) algorithm, although Bayesian computation carried out via the Gibbs sampler is also very promising. In this paper I present the maximum-likelihood-based approach.

McLachlan and Peel, 2000 suggest the following interpretation of finite mixture models. The available data of T realizations $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_T$ of the sequence of the innovation vectors $\{\boldsymbol{\varepsilon}_t\}$ are viewed as being incomplete, since their associate component-indicator vectors, $\boldsymbol{z}_1, \dots, \boldsymbol{z}_T$, are not available. This approach could also be used to fit mixture models in the Bayesian framework. The use of component-label vectors is one of the obvious ways to interpret mixture models. For convenience, a G-dimensional component-label vector $\boldsymbol{z}_t$ could be introduced, where the gth element of $\boldsymbol{z}_t$, z_{gt} is defined to be one or zero, according to whether the component of origin of $\boldsymbol{\varepsilon}_t$ in the mixture is equal to g or not.

The component-label vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ are taken to be the T realized values of the latent random sequence $\{\boldsymbol{z}_t\}$. The gth mixing proportion λ_g can be viewed as the prior probability that the entity belongs to the gth component of the mixture. Applying Bayes theorem, the posterior probability that the entity belongs to the gth component given that $\boldsymbol{\varepsilon}_t$ has been observed, is

$$\text{Prob}\{\mathbf{z}_{gt} = 1 | \boldsymbol{\varepsilon}_t\} = \frac{\lambda_g \phi(\boldsymbol{\varepsilon}_t; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{g,t})}{f(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi})}. \tag{12}$$

The EM algorithm exploits the simplified form of log likelihood function due to the assumption of the incomplete data. The log likelihood function now could be written as the sum of the log likelihood functions across all components, which makes its treatment easier. The EM algorithm overcomes the fact that the component-indicator vectors $\boldsymbol{z}_t$ are unknown by iteratively working with the conditional expectation of the complete-data log likelihood.

The log likelihood function of unknown parameters $\boldsymbol{\Psi}$ using the latent variables $\boldsymbol{z}_t$ could be written as

$$\mathcal{L}(\boldsymbol{\Psi}) = \sum_{g=1}^{G} \sum_{t=1}^{T} z_{gt} \left(\log \lambda_g + \log \phi(\boldsymbol{\varepsilon}_t; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{g,t}) \right). \tag{13}$$

Recalling the well-known form of the function $\phi(\cdot)$ and using the representation in (3) one can re-write the log likelihood function in (13) (up to constants and scaling) as

$$\mathcal{L}(\boldsymbol{\Psi}) = \sum_{g=1}^{G} \sum_{t=1}^{T} z_{gt} \mathcal{L}(g,t), \tag{14}$$

$$\mathcal{L}(g,t) = 2\log \lambda_g - 2\log |\boldsymbol{D}_{g,t}| - \log |\boldsymbol{C}_{g,t}| - (\boldsymbol{\varepsilon}_t - \boldsymbol{\mu}_g)' \boldsymbol{D}_{g,t}^{-1} \boldsymbol{C}_{g,t}^{-1} \boldsymbol{D}_{g,t}^{-1} (\boldsymbol{\varepsilon}_t - \boldsymbol{\mu}_g). \tag{15}$$

The EM algorithm, first suggested by Dempster et al., 1977, consists of an E-step and a M-step, which iteratively calculate $\mathcal{L}(\boldsymbol{\Psi}^{(k)})$. The steps are alternated repeatedly until the difference $\mathcal{L}(\boldsymbol{\Psi}^{(k+1)}) - \mathcal{L}(\boldsymbol{\Psi}^{(k)})$ changes by an arbitrarily small amount. Wu, 1983 proved the main properties of the EM algorithm under assumptions of compactness of the domain of $\mathcal{L}(\boldsymbol{\Psi})$ and continuity and differentiability of $\mathcal{L}(\boldsymbol{\Psi})$: (1) if $\mathcal{L}(\boldsymbol{\Psi})$ is bounded, $\boldsymbol{\Psi}^{(k)}$ converges to a stationary point $\boldsymbol{\Psi}^*$; (2) If all stationary points are local maxima and this set is discrete, then $\boldsymbol{\Psi}^{(k)}$ converges to a local maximum; (3) If $\mathcal{L}(\boldsymbol{\Psi})$ is unimodal, then $\boldsymbol{\Psi}^{(k)}$ converges to the global maximum. Therefore, to escape local maxima, one could apply such approaches as random restart or simulated annealing methods. To avoid singularities of components, a prior distribution of the parameters can be added or the parameter set could be restricted.

The (k+1)th step of the estimation algorithm is the following.

E-Step (Expectation)

The E-Step handles the added unobservable data $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ by taking the conditional expectation $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ of the complete-data log likelihood given the observed data $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T]$ and using estimates of the unknown set of parameters $\boldsymbol{\Psi}^{(k)}$ attained on the previous iteration:

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = \mathbb{E}\left(\mathcal{L}(\boldsymbol{\Psi}) | \{\boldsymbol{\varepsilon}_t\}, \boldsymbol{\Psi}^{(k)}\right) = \sum_{g=1}^{G} \sum_{t=1}^{T} \mathbb{E}\left(z_{gt} | \boldsymbol{\varepsilon}_t, \boldsymbol{\Psi}^{(k)}\right) \mathcal{L}(g,t). \tag{16}$$

It follows from (12) that

$$\mathbb{E}\left(z_{gt} | \boldsymbol{\varepsilon}_t, \boldsymbol{\Psi}^{(k)}\right) = \frac{\lambda_g^{(k)} \phi(\boldsymbol{\varepsilon}_t; \boldsymbol{\mu}_g^{(k)}, \boldsymbol{\Sigma}_{g,t}^{(k)})}{f(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})}, \tag{17}$$

therefore the conditional expectation is

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = \sum_{g=1}^{G} \sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}) \mathcal{L}(g,t), \tag{18}$$

where $\tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})$ replaces the estimated $\mathbb{E}\left(z_{gt} | \boldsymbol{\varepsilon}_t, \boldsymbol{\Psi}^{(k)}\right)$ for simplicity.

M-Step (Maximization)

The M-Step gives the updated estimate $\boldsymbol{\Psi}^{(k+1)}$ by maximizing $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ with respect to $\boldsymbol{\Psi}$ over the parameter space.

It follows directly from (18) and (15) that the updated estimate of the mixing proportions is

$$\hat{\lambda}_g^{(k+1)} = \frac{1}{T} \sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}), \quad g = 1, \ldots, G, \tag{19}$$

because they do not change with time.

It is possible to get estimates of other parameters by the following multi-step procedure.

First, one sets $C_{g,t}$ to be equal to the $n \times n$ identity matrix (see Engle and Sheppard, 2001 for more details), then the conditional expectation from the E-Step (ignoring constant terms) becomes

$$
\begin{aligned}
Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) &= -\sum_{g=1}^{G} \sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}) \left(2 \log |\boldsymbol{D}_{g,t}| + (\boldsymbol{\varepsilon}_t - \boldsymbol{\mu}_g)' \boldsymbol{D}_{g,t}^{-1} \boldsymbol{D}_{g,t}^{-1} (\boldsymbol{\varepsilon}_t - \boldsymbol{\mu}_g)\right) \\
&= -\sum_{g=1}^{G} \sum_{i=1}^{n} \sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}) \left(2 \log(\boldsymbol{D}_{g,ii,t}) + \frac{(\boldsymbol{\varepsilon}_{i,t} - \mu_{g,i})^2}{\boldsymbol{D}_{g,ii,t}^2}\right).
\end{aligned} \tag{20}
$$

Maximizing this function $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ is equivalent to estimating $n \times G$ independent GARCH models with a weighting scheme given by $\tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})$. However, there is an additional constraint in the form of $\sum_{g=1}^{G} \lambda_g \boldsymbol{\mu}_g = \mathbf{0}$. To avoid the joint estimation of $4 \times G \times n$ parameters, one can apply a two step procedure. First, construct a consistent estimator of $\boldsymbol{\mu}_g$, e.g., taking $\boldsymbol{D}_{g,ii,t}$ to be constant across time and maximizing $Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)})$ with respect to $\boldsymbol{\mu}_g$. It leads to the following estimator:

$$
\hat{\boldsymbol{\mu}}_g^{(k+1)} = \frac{\sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}) \boldsymbol{\varepsilon}_t}{\sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})}, \quad g = 1, \ldots, G. \tag{21}
$$

Then, conditioning on $\hat{\boldsymbol{\mu}}_g^{(k+1)}$, $G \times n$ independent univariate GARCH models could be estimated using the series $\{\varepsilon_{i,t}\}$ and the weighting scheme $\tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})$ for the corresponding log likelihood functions.

Given estimates of $\boldsymbol{D}_{g,ii,t}$, it is possible to reestimate $\hat{\boldsymbol{\mu}}_g^{(k+1)}$ taking into account the dynamics of the variances. Denote the "variance corrected" weighting scheme

$$
\tilde{\tau}_{g,i,t} = \frac{\tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})}{\hat{\boldsymbol{D}}_{g,ii,t}^2}. \tag{22}
$$

Then $(G+1) \times n$ first order conditions derived from the loglikelihood in (20) with respect to $\mu_{g,i}$ are:

$$
\sum_{t=1}^{T} \tilde{\tau}_{g,i,t}(\boldsymbol{\varepsilon}_{i,t} - \mu_{g,i}) + \eta \lambda_g = 0, \ \sum_{g=1}^{G} \lambda_g \mu_{g,i} = 0, \ g = 1, \ldots, G, \ i = 1, \ldots, n. \tag{23}
$$

The solution is

$$\hat{\mu}_{g,i}^{(k+1)} = \frac{\sum_{t=1}^{T} \tilde{\tau}_{g,i,t}\boldsymbol{\varepsilon}_{i,t}}{\sum_{t=1}^{T} \tilde{\tau}_{g,i,t}} + \eta \upsilon_g, \quad \eta = -\frac{\sum_{g=1}^{G}\sum_{t=1}^{T} \upsilon_g \tilde{\tau}_{g,i,t}\boldsymbol{\varepsilon}_{i,t}}{\sum_{g=1}^{G} \upsilon_g \hat{\lambda}_g}, \quad \upsilon_g = \frac{\hat{\lambda}_g}{\sum_{t=1}^{T} \tilde{\tau}_{g,i,t}}. \tag{24}$$

After estimating all the parameters of univariate GARCH models, one has to calculate standardized residuals $\hat{\boldsymbol{\eta}}_{g,t}$ for each matrix $\hat{\boldsymbol{D}}_{g,t}$ and G estimates of the unconditional correlation matrices $\bar{\boldsymbol{Q}}_g^{(k+1)}$ as described in Aielli, 2011, using weights $\tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})$ instead of $\frac{1}{T}$ in all the formulae. If we replace the matrices $\boldsymbol{D}_{g,t}$ in (18) by their estimates and ignore constant terms, the conditional likelihood now is given by

$$Q(\boldsymbol{\Psi}; \boldsymbol{\Psi}^{(k)}) = -\sum_{g=1}^{G}\sum_{t=1}^{T} \tau_g(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)}) \left(\log |\boldsymbol{C}_{g,t}| + \hat{\boldsymbol{\eta}}_t' \boldsymbol{C}_{g,t}^{-1} \hat{\boldsymbol{\eta}}_t\right). \tag{25}$$

Estimates of the parameters of G processes describing correlation dynamics for each component could be found by solving G independent maximization problems of CCC, DCC and DECO estimation procedures (see corresponding papers), but putting weights $\tau_j(\boldsymbol{\varepsilon}_t; \boldsymbol{\Psi}^{(k)})$ on the t-th summand of the conditional likelihood for each g, $g = 1, \ldots, G$.

4 Financial Application

Under the assumption of model (2), it is not difficult to prove that the random variable, which corresponds to the return of a portfolio of assets with return vector $\boldsymbol{r}_t$, conditionally follows a univariate finite mixture distribution. The proof based on properties of characteristic functions can be found in Paolella, 2011. Taking into account the notation in the previous section, the result can be written as follows.

Let $\mathcal{P}_t = \boldsymbol{w}'\boldsymbol{r}_t$, $\boldsymbol{w}'\mathbf{1} = 1$ be the return of a portfolio with weights given by a non-random vector $\boldsymbol{w}$. Then the density of the distribution of $\mathcal{P}_t|\Omega_{t-1}$ is described by the following function:

$$f_{\mathcal{P}}(x) = \sum_{g=1}^{G} \lambda_g \phi(\boldsymbol{x}; \boldsymbol{w}'(\boldsymbol{\mu}_t + \boldsymbol{\mu}_g), \boldsymbol{w}'\boldsymbol{\Sigma}_{g,t}\boldsymbol{w}), \quad \lambda_g > 0 \ \forall g, \quad \sum_{g=1}^{G} \lambda_g = 1, \quad \sum_{g=1}^{G} \lambda_g \boldsymbol{\mu}_g = \mathbf{0}. \tag{26}$$

Estimation of the multivariate normal conditional correlation model allows us to predict the density $f_{\mathcal{P}}(x)$ one step ahead by predicting $\boldsymbol{\Sigma}_{g,t}$ for every component and $\boldsymbol{\mu}_t$, since other parameters are time-invariant in the model. The predicted density of the portfolio return is then (with the same constraints)

$$\hat{f}_{\mathcal{P},t+1}(x) = \sum_{g=1}^{G} \lambda_g \phi(\boldsymbol{x}; \boldsymbol{w}'(\hat{\boldsymbol{\mu}}_{t+1} + \boldsymbol{\mu}_g), \boldsymbol{w}'\hat{\boldsymbol{\Sigma}}_{g,t+1}\boldsymbol{w}). \tag{27}$$

Using the predicted density, we can derive a number of useful risk measures. For simplicity I write $\mathcal{P}_{t+1}$ instead of $\mathcal{P}_{t+1}|\Omega_t$ in all the formulae below and use the following notation: $\hat{\mu}^{\mathcal{P}}_{g,t+1} = \boldsymbol{w}'(\boldsymbol{\mu}_t + \boldsymbol{\mu}_g)$ and $\hat{\sigma}^{2,\mathcal{P}}_{g,t+1} = \boldsymbol{w}'\hat{\boldsymbol{\Sigma}}_{g,t+1}\boldsymbol{w}$. $\Phi(x, \mu, \sigma^2)$ denotes the cdf of the Gaussian random variable with the mean μ and the variance σ^2 evaluated at a point x.

1. Variance

The variance is the most popular risk measure. Since the seminal work of Markowitz, 1952 it is used for portfolio selection in the mean-variance framework.

$$\mathbb{V}[\mathcal{P}_{t+1}] = \sum_{g=1}^{G} \lambda_g \left(\hat{\sigma}^{2,\mathcal{P}}_{g,t+1} + (\hat{\mu}^{\mathcal{P}}_{g,t+1})^2\right) - \left(\sum_{g=1}^{G} \lambda_g \hat{\mu}^{\mathcal{P}}_{g,t+1}\right)^2. \tag{28}$$

2. Mean absolute deviation

The mean absolute deviation can be used instead of the variance. This measure is more robust to outliers, since all distances from the mean are treated equally, whereas for the variance large distances are amplified even more.

$$\begin{aligned}
\mathbb{E}[|\mathcal{P}_{t+1} - \mathbb{E}[\mathcal{P}_{t+1}]|] &= \sum_{g=1}^{G} \lambda_g \left((P - \hat{\mu}^{\mathcal{P}}_{g,t+1})\mathrm{erf}\left(\frac{P - \hat{\mu}^{\mathcal{P}}_{g,t+1}}{\sqrt{2}\hat{\sigma}^{2,\mathcal{P}}_{g,t+1}}\right)\right) \\
&+ \sum_{g=1}^{G} \lambda_g \left(2\hat{\sigma}^{2,\mathcal{P}}_{g,t+1}\phi(P, \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1})\right), \qquad (29) \\
P = \mathbb{E}[\mathcal{P}_{t+1}] &= \sum_{g=1}^{G} \lambda_g \hat{\mu}^{\mathcal{P}}_{g,t+1}. \qquad (30)
\end{aligned}$$

3. Lower partial moments

The measure of downside risk somewhat alternative to the standard deviation is the lower partial moment. It means that one take into account only negative deviations from a threshold (often taken to be zero) and ignores positive deviations. It is reasonable from the point of view that an investor is not against large positive returns, but large negative returns are undesirable. This measure can be used to derive CAPM (Bawa and Lindenberg, 1977), to price assets

(Harlow and Rao, 1989), to evaluate investment performance (Sortino and Price, 1994) and to choose an optimal portfolio (Jarrow and Zhao, 2006).

The first-order lower partial moment with respect to a threshold $\bar{\mathcal{P}}$ is

$$\mathbb{E}[\max(\bar{\mathcal{P}} - \mathcal{P}_{t+1}, 0)] = \sum_{g=1}^{G} \lambda_g [(\bar{\mathcal{P}} - \hat{\mu}^{\mathcal{P}}_{g,t+1})\Phi(\bar{\mathcal{P}}, \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}) + \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}\phi(\bar{\mathcal{P}}, \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1})]. \quad (31)$$

The second-order lower partial moment with respect to a threshold $\bar{\mathcal{P}}$ is

$$\begin{aligned} \mathbb{E}[\max(\bar{\mathcal{P}} - \mathcal{P}_{t+1}, 0)^2] &= \sum_{g=1}^{G} \lambda_g \left((\hat{\sigma}^{2,\mathcal{P}}_{g,t+1} + (\bar{\mathcal{P}} - \hat{\mu}^{\mathcal{P}}_{g,t+1})^2)\Phi(\bar{\mathcal{P}}, \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}) \right) \\ &+ \sum_{g=1}^{G} \lambda_g \left((\bar{\mathcal{P}} - \hat{\mu}^{\mathcal{P}}_{g,t+1})\hat{\sigma}^{2,\mathcal{P}}_{g,t+1}\phi(\bar{\mathcal{P}}, \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}) \right). \quad (32) \end{aligned}$$

4. Expected Shortfall

Artzner et al., 1997 introduce expected shortfall as one of the coherent risk measures satisfying properties of monotonicity, sub-additivity, homogeneity, and translational invariance. The expected shortfall is defined to be the expected loss of portfolio value given that a loss is occurring at or below the γ-quantile. The practical application of this risk measure is demonstrated in Bertsimas et al, 2004. The formula of the expected shortfall of the finite mixture of normal distributions is derived by Paolella, 2011.

$$\mathrm{ES}_\gamma(\mathcal{P}_{t+1}) = \frac{1}{\gamma}\sum_{g=1}^{G} \lambda_g \left(\hat{\mu}^{\mathcal{P}}_{g,t+1}\Phi(Q_{\mathcal{P}}(\gamma), \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}) - \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}\phi(Q_{\mathcal{P}}(\gamma), \hat{\mu}^{\mathcal{P}}_{g,t+1}, \hat{\sigma}^{2,\mathcal{P}}_{g,t+1}) \right), \quad (33)$$

where $Q_{\mathcal{P}}(\gamma)$ is the γ-quantile of $\mathcal{P}_{t+1}$.

5. Rachev Ratio

The Rachev ratio, which is the ratio of the expected tail gains to the expected tail losses, is presented in Biglova et al., 2004. Values γ_1 and γ_2 denote the order of quantiles of $\mathcal{P}_{t+1}$. An example of application of this ratio to the portfolio selection could be found in Satchell, 2010.

$$RR_{\gamma_1,\gamma_2}(\mathcal{P}_{t+1}) = \frac{\sum_{g=1}^{G} \lambda_g \hat{\mu}^{\mathcal{P}}_{g,t+1} - \gamma_2 \mathrm{ES}_{\gamma_2}(\mathcal{P}_{t+1})}{(1-\gamma_2)\mathrm{ES}_{\gamma_1}(\mathcal{P}_{t+1})} \quad (34)$$

5 Data

The considered data consist of 30 stock return time series. The source of the data is the CRSP database. The frequency is daily and there are 5295 observations in total, which correspond to 21 years (from January 1, 1990 till December 31, 2010). On January 1, 2012 27 of these thirty stocks were S&P 500 components.[1] These thirty stocks are the first thirty in a larger database, where securities are ordered by their PERMNO.[2] Since the stocks are not sorted out by any characteristics of their returns, we can believe that they are randomly selected for the analysis.

Five data sets are made up of these 30 stocks. Data Sets 1 to 3 include stocks #1 to #5, #6 to #10 and #11 to #15 correspondingly. Data Set 4 includes stocks #6 to #15 and Data Set 5 includes stocks #16 to #30.

I pursue the analysis and comparisons of different models using a "rolling windows" technique. I fix the window size to be 500 observations which approximately corresponds to two business years. After I have made some estimations using only the information from a certain window, I move the window one day ahead, "forgetting" the first observation of the previous window and "learning" a new piece of information. The time span of the data results in 4795 windows.[3]

Table 1 displays names of the companies in the data and some useful statistical properties of their stock returns. All the mean returns are positive and only for 2 companies the hypothesis that the mean return is equal to zero is not rejected at least at the 10% significance level.[4] The significant estimates of the mean are heterogeneous, varying from 0.0410% (Consolidated Edison) to 0.1679 % (EMC Group) per day. If we consider the Sharpe ratios of the stocks, we see that 6 stocks have the Sharpe ratio between 0.02 and 0.03, 9 stocks have the Sharpe ratio between 0.03 and 0.04 and 11 stocks have the Sharpe ratio between 0.04 and 0.05. Only one stock, excluded from the index at the end of the time span, has the very low Sharpe ratio of 0.004.

[1] Fortune Brands (#7) was removed on October 4, 2011 due to the spinoff of Beam Inc. Allegheny Energy (#3) was removed on February 25, 2011, because it was acquired by First Energy. Eastman Kodak (#20) was removed on December 17, 2010 due to market capitalization changes.

[2] PERMNO is a unique permanent security identification number assigned by CRSP to each security. It neither changes during an issue's trading history, nor is it reassigned after an issue ceases trading.

[3] I must have at least one observation left after the window to evaluate models' forecasting ability.

[4] Note that I use the heteroscedasticity and autocorrelation-consistent estimate of the variance for tests in this table. Crack and Ledoit, 2010 explain why it is so important to account for possible dependences in the financial data while applying the central limit theorem.

Table 1: Stock names and basic statistical properties of their returns

No.	Name	Mean, %	S.D.,%	Sharpe Ratio	AR(1)	Fraction of signif., %
1	Oracle Corp	***0.1317	3.0989	0.0425	***-0.0425	30.19
2	Microsoft Corp	***0.1004	2.0626	0.0487	-0.0200	25.29
3	Allegheny Energy	0.0431	2.2134	0.0195	*0.0241	24.29
4	T. Rowe Price Group	***0.1101	2.3901	0.0461	-0.0051	43.04
5	Honeywell Intl	**0.0652	2.0390	0.0320	0.0038	17.08
6	EMC Corp	***0.1679	3.1446	0.0534	0.0116	11.93
7	Fortune Brands	*0.0470	1.8201	0.0258	0.0197	10.53
8	Linear Technology Corp	***0.1296	2.5428	0.0510	**-0.0335	13.76
9	Archer Daniels Midland	**0.0491	1.7535	0.0280	***-0.0794	46.50
10	Fiserv	***0.0921	1.9088	0.0483	***-0.1100	49.96
11	Cerner Corp	***0.1520	3.1542	0.0482	***-0.0530	19.68
12	Dell	***0.1548	3.1391	0.0493	0.0211	2.67
13	Coca-Cola	***0.0555	1.4418	0.0385	0.0018	13.03
14	Consolidated Edison	***0.0410	1.1542	0.0355	-0.0193	22.29
15	Celgene Corp	***0.1628	3.8146	0.0427	0.0038	15.16
16	DENTSPLY Intl	***0.0891	1.8443	0.0483	***-0.0521	34.09
17	Fastenal	***0.1129	2.4604	0.0459	***0.0418	18.49
18	DTE Energy	**0.0422	1.2023	0.0351	***-0.0588	32.28
19	E. I. du Pont de Nemours	**0.0471	1.6789	0.0281	*-0.0231	10.97
20	Eastman Kodak	0.0104	2.5562	0.0040	**0.0342	17.43
21	Eaton Corp	***0.0659	1.7475	0.0377	***-0.0383	26.92
22	Exxon Mobil Corp	***0.0572	1.1535	0.0496	***-0.1145	49.33
23	Waste Management	***0.0916	2.4335	0.0376	-0.0013	27.69
24	BMC Software	***0.1085	3.0322	0.0358	*0.0258	31.73
25	General Dynamics Corp	***0.0816	1.7735	0.0460	0.0071	34.05
26	General Electric	**0.0512	1.7405	0.0294	-0.0083	4.65
27	Lab. Corp of American H.	*0.0665	2.5732	0.0259	***-0.0613	19.81
28	Novellus Systems	***0.1377	3.5985	0.0383	0.0022	19.58
29	People's United Financial	***0.0944	2.3775	0.0397	***-0.1059	42.20
30	Goodrich Corp	**0.0623	2.1057	0.0296	**0.0301	39.78

Note: Means and standard deviations are expressed on the daily basis and in percent. "Fraction of signif." stays for the number of the rolling windows, where the AR(1) coefficient is significant at the 5% level divided by the total number of these windows. ***, ** and * indicate the significance levels of 1%, 5% and 10% correspondingly.

The analysis of the AR(1) coefficient of the individual return time series shows that for 17 of 30 stocks the coefficient is significantly different from zero at least at the 10% confidence level and for 11 of these 17 even at the 1% confidence level. Negative AR(1) coefficients prevail among the significant AR(1) coefficients. If we consider the fraction of the rolling windows where AR(1) coefficients are significant, we see that for 11 stocks this fraction is higher than 30 % and for two of them is close to 50% with the coefficient value of approximately -0.11. Arguments for the existence of the autocorrelation in the daily data could be found in Brown and Warner, 1985 and in Gu and Finnerty, 2002. Sentana and Wadhwani, 1994 explain how higher serial correlation is still compatible with the equilibrium. Due to the link between the volatility and the autocorrelation, which they detect, higher predictability of returns will be outweighed by a higher risk. Therefore, autocorrelation would be worth taking into account while modeling the conditional mean vector.

Table 2 addresses the issue of the number of the components in the mixture. On the one hand, we can simply rely on the information from the data and select the appropriate number of the components relying on the AIC / BIC information criteria. On the other hand, since we deal with the financial data, we can appeal to some empirically observed facts of the market behavior. One can, for example, assign a two component mixture to bull and bear markets or a three component mixture to a stable market, a boom phase and a recession. Explaining four and more components is a challenging task.

To investigate this problem, I fit to each univariate return time series in the data set a three component univariate normal mixture model. For 3 stocks the EM-algorithm does not converge. Under the assumption that the EM algorithm indeed converges to the ML global maximum, the estimated parameters are asymptotically normally distributed with the covariance matrix being equal to the inverted Fisher information matrix.[5] This allows for some tests. However, under the null hypothesis that the smallest component probability λ_3 is zero (or the number of components is equal to two), the distribution of λ_3 is not normal, because, first, its value lies on the boundary, second, the mean and the variance of the third component are undefined.

The obvious null hypothesis for all the component means μ_g will be H_0: $\mu_g = 0, g = 1, 2, 3$ (individually). This allows for understanding whether there is a boom with positive returns, a recession with negative returns or a stable phase where returns are zero on average. According to the results of these tests, we can identify two groups of stocks. The one group of 19 stocks consists of stocks for which only one μ_g is significantly different from zero at least at the 10% level. For 12 stocks this is μ_1, for 3 stocks μ_2 and for 4 stocks μ_3. These means are all except that of stock #1 positive varying from -0.11% to 2.17%, probably indicating periods of high returns. In most of the cases these high returns are accompanied by moderate or high volatility (see columns with the values of σ_g). The other group of 6 stocks includes stocks with at least two significant μ_g, and these μ_g also have opposite signs. In all the cases these are of components 1 and 2 indicating that even if there are zero returns periods, their probability is small and the most significant driving forces for these assets are "bearish" and "bullish" trends. The volatility of the component with the negative significant mean is always the lowest, meaning that the level of negative returns is more stable than that of positive returns. I also tested for

[5] To make tests robust to density misspecifications, I apply the so called "sandwich" form of the covariance matrix estimator: the Fisher information matrix multiplied on both sides by the inverted Hessian matrices of the loglikelihood function.

the H_0: $\sigma_1 = \sigma_2 = \sigma_3$ and it was rejected at the 1% significance level for all stocks.

Table 2: Estimated parameters of a three component univariate mixture normal distribution

H_0				$\mu_1 = 0$	$\mu_2 = 0$	$\mu_3 = 0$				G=2	G=3
No.	λ_1	λ_2	λ_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	p-value	p-value
1	0.54	0.43	0.03	**-0.11	0.33	1.58	1.73	3.77	10.24	0.017	0.007
2	0.63	0.31	0.06	**0.17	*-0.10	0.38	2.17	0.90	5.14	0.029	0.009
3	0.61	0.37	0.02	*0.04	0.08	-0.33	0.80	2.15	9.50	0.007	0.003
4	0.49	0.38	0.13	-0.07	0.00	1.13	2.50	1.13	5.30	0.032	0.556
5	0.62	0.36	0.02	0.01	*0.17	-0.10	1.19	2.63	7.70	0.023	0.109
6	0.53	0.44	0.03	-0.15	0.52	0.71	1.79	4.00	9.12	0.017	0.033
7	0.49	0.41	0.10	***0.15	-0.07	0.04	1.62	0.84	3.74	0.028	0.247
8	0.47	0.36	0.17	-0.10	-0.18	***1.41	1.54	3.40	5.22	0.033	0.450
9	0.57	0.38	0.05	*0.16	-0.13	0.13	1.96	1.00	5.76	0.013	0.009
10	0.51	0.44	0.05	0.05	0.01	***1.22	1.22	2.64	5.55	0.029	0.042
11	0.53	0.45	0.02	0.00	**0.23	2.10	1.61	3.97	12.57	0.024	0.079
12	0.57	0.37	0.06	**0.34	**-0.14	0.15	3.31	1.50	7.24	0.026	0.067
13	0.56	0.38	0.06	*0.08	-0.03	0.35	1.57	0.74	3.74	0.022	0.757
14	0.48	0.46	0.06	*0.08	-0.01	0.20	0.77	1.28	2.82	0.038	0.282
15	0.55	0.31	0.14	-0.07	0.04	***1.37	3.45	1.52	7.58	0.014	0.551
16	Algorithm does not converge										
17	0.59	0.32	0.09	**0.17	-0.05	0.35	2.33	1.07	5.61	0.021	0.725
18	0.52	0.45	0.03	0.04	**0.06	-0.09	1.38	0.71	3.82	0.029	0.037
19	0.63	0.31	0.06	*-0.06	**0.26	0.04	1.14	2.22	4.14	0.029	0.541
20	0.58	0.37	0.05	**-0.09	*0.16	0.13	1.17	2.55	8.30	0.019	0.176
21	0.53	0.42	0.05	**0.12	-0.04	0.38	1.81	0.94	4.47	0.020	0.461
22	0.64	0.35	0.01	**0.07	0.01	0.75	1.01	1.90	6.35	0.019	0.481
23	0.60	0.38	0.02	0.12	-0.08	**2.17	2.42	0.79	9.30	0.017	N/A
24	0.51	0.48	0.01	***0.28	-0.05	-0.98	3.79	1.49	12.98	0.020	0.007
25	0.54	0.40	0.06	***0.12	-0.02	0.41	1.75	0.79	4.53	0.022	0.204
26	0.50	0.43	0.07	***0.16	-0.05	-0.05	1.76	0.86	4.77	0.033	0.044
27	Algorithm does not converge										
28	0.58	0.40	0.02	**0.28	***-0.24	**3.23	4.37	1.88	10.07	0.030	0.069
29	Algorithm does not converge										
30	0.57	0.37	0.06	*0.13	**-0.09	0.30	2.00	0.86	4.98	0.029	0.066

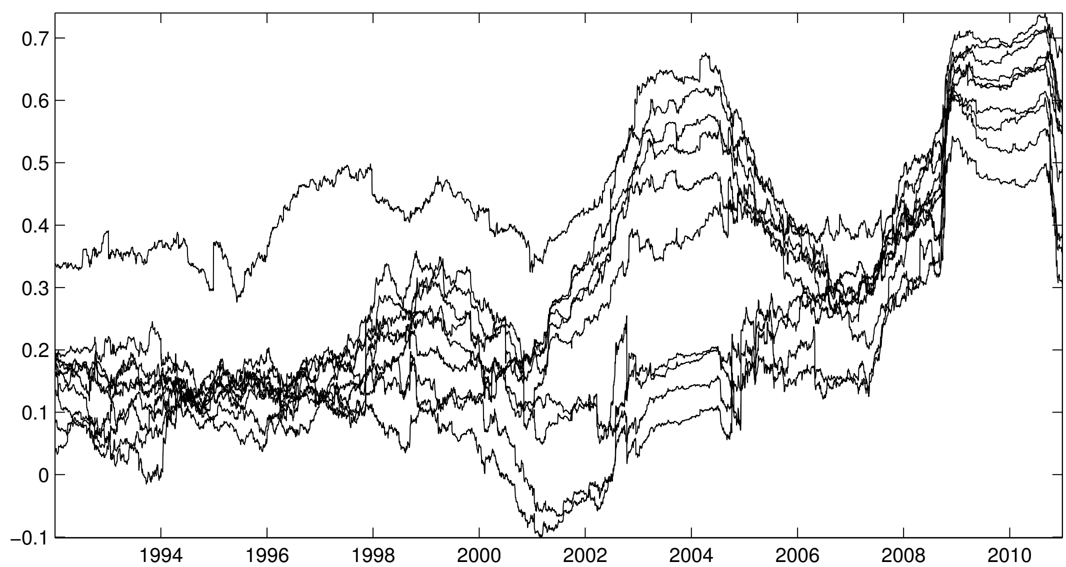
Note: I fit a three component univariate mixture normal distribution to the stock returns from #1 to #30. The table shows the parameter estimates and significance of the corresponding statistics / p-values of the following hypotheses. (1) H_0: $\mu_g = 0$, $g = 1, 2, 3$ versus H_1: $\mu_g \neq 0$, $g = 1, 2, 3$; (2) H_0: $G = 2$ versus H_1: $G = 3$; (3) H_0: $G = 3$ versus H_1: $G = 4$. ***, ** and * indicate the significance levels of 1%, 5% and 10% correspondingly. Two last columns display p-values for hypotheses (2) and (3).

To test the hypothesis concerning the number of components for each of 30 stocks, I use the parametric bootstrapping of the loglikelihood ratio. This method is suggested in McFadden, 1987. The algorithm is the following (1) Estimate the parameters assuming a G-component mixture of normal distributions; (2) Use these parameters to simulate from this distribution: I draw 1000 samples of the length of the original sample; (3) Use these samples (including the original one) to estimate the parameters of a $(G+1)$-component mixture of normal distributions (the alternative hypothesis); (4) Calculate the loglikelihood ratio statistics for each sample and an empirical p-value. If this value is smaller than some predefined significance level, this would mean that we cannot reject the null hypothesis.

For $H_0 : G = 1$ the empirical p-value is zero for all the stocks meaning that there is a statistical evidence for a better fit of a two component mixture of normal distributions. In Table 2, the two last columns present the corresponding values of the empirical p-value for $H_0 : G = 2$ and $H_0 : G = 3$. We see that the null hypothesis $H_0 : G = 2$ is rejected for 26 stocks at the 5% level (for 3 stocks the algorithm does not converge meaning that at most a two-component mixture could be fitted), but cannot be rejected at the 1% level (except for stock #3). The null hypothesis $H_0 : G = 3$ is rejected for 5 stocks at the 1% level, but cannot be rejected for 13 stocks even at the 10% level. Besides there are many cases when the EM-algorithm does not converge while doing these tests. Taking into account that empirical analysis in the next section is done for multivariate models and using about ten times less observations in the rolling window, I decide for the two-component mixture to ensure identification.

Since my family of models treats variances and correlations separately, it is interesting to look at the roughly estimated dynamics of correlations over time. To do so, I take Data Set 1 where we have 10 different stock pairs and estimate for them unconditional correlations on the rolling window of 500 observations (it makes the graphs smooth, but they are not necessary so).

Figure 1: Empirical correlations, an example



Note: Pairwise correlations are estimated for Data Set 1 using the rolling window of 500 observations.

Figure 1 shows that correlations are indeed time-varying: some of the pairwise correlations were close to -0.1 in 2001 and some of them were close to 0.7 in 2010. The DCC-based model is expected to work especially well in the period from 1999 to 2001 when all correlations simultaneously decreased and in the period from 2007 to 2009 when the correlations rose. The DECO-based model may face some fitting troubles in the period from 2001 to 2005, when the values of pairwise correlations are visually very different from each other, with four of them being closer to 0.1 and six of them lying around 0.5.

6 Empirical Application

This chapter consists of three parts. In the first part I analyse the parameter estimates of all three suggested model specifications and test four types of hypotheses related to differences in parameter values. Then I examine the time-varying behaviour of essential parts of the model, such as the component probability, the component mean, the component variance and correlations. Finally, I consider the predictive ability of the model family and compare it to that of its feasible competitors using three different tests.

To each data set I fit multivariate normal conditional correlation models with CCC, DCC and DECO processes applied to describe the dynamics of the correlation matrices. I denote these models "MixCCC", "MixDCC" and "MixDECO". GARCH processes are GARCH(1,1) in all the cases and the dynamic correlation processes are DCC(1,1) and DECO(1,1). These are specifications, which are the most common in the literature. For all stock returns I do AR(1) pre-filtering in accordance with the findings in the previous section. Since even for the whole data set of 5295 observation a three-component mixture of normal distributions is not a proper model for some of the stocks, I choose a two-component mixture model to fit the data of 500 observation windows. It allows for avoiding nonidentifiability due to overfitting, which, according to Crawford, 1994, is more problematic than its other causes.

Table 3 shows the statistics of GARCH parameter estimates of the first three stocks of Data Set 1. The mean and the standard deviation are calculated over estimates made on 4795 windows. Since I do not assume asymptotic normality, I do not draw any conclusion about significance of these estimates, they are just those providing the best fits in terms of likelihood maximization. Table 4 shows the statistics of the estimates of component probability and the correlation process parameters for all five data sets across all the models. One can see that the average values of parameter λ are similar across data sets and lie around 0.8.

Table 3: Statistics of the GARCH parameter estimates across the models

	ω_1	α_1	$\alpha_1+\beta_1$	ω_2	α_2	$\alpha_2+\beta_2$
The mixture CCC-GARCH(1,1)						
Oracle Corp						
Mean	0.0001	0.0473	0.7671	0.0015	0.4951	0.8480
S.D.	0.0002	0.0499	0.2710	0.0018	0.3983	0.2431
Microsoft Corp						
Mean	0.0000	0.0748	0.8680	0.0002	0.1462	0.8916
S.D.	0.0001	0.0537	0.1799	0.0003	0.1818	0.2007
Allegheny Energy						
Mean	0.0000	0.0852	0.9409	0.0002	0.2584	0.9172
S.D.	0.0000	0.0533	0.0743	0.0007	0.3405	0.1535
The mixture DCC(1,1)-GARCH(1,1)						
Oracle Corp						
Mean	0.0002	0.0460	0.7569	0.0016	0.5077	0.8543
S.D.	0.0002	0.0472	0.2791	0.0018	0.3988	0.2386
Microsoft Corp						
Mean	0.0000	0.0733	0.8719	0.0001	0.1476	0.9093
S.D.	0.0001	0.0533	0.1696	0.0002	0.1677	0.1877
Allegheny Energy						
Mean	0.0000	0.0837	0.9437	0.0002	0.2680	0.9160
S.D.	0.0000	0.0518	0.0600	0.0006	0.3481	0.1594
The mixture DECO(1,1)-GARCH(1,1)						
Oracle Corp						
Mean	0.0001	0.0524	0.7770	0.0014	0.4109	0.8153
S.D.	0.0002	0.0613	0.2713	0.0016	0.4078	0.2782
Microsoft Corp						
Mean	0.0000	0.0700	0.8682	0.0002	0.1486	0.8814
S.D.	0.0001	0.0583	0.1850	0.0003	0.1794	0.1933
Allegheny Energy						
Mean	0.0000	0.0913	0.9322	0.0001	0.2213	0.9055
S.D.	0.0001	0.0915	0.1092	0.0002	0.3022	0.1488

Note: Summary of the estimates of the GARCH-based part is presented for three types of the correlation matrix dynamics and for the first three stocks in the sample. The mean and the standard deviation of GARCH parameters are calculated on the base of model estimates on 4795 rolling windows.

It is interesting and important to test the hypothesis about the parameter equality across the components and across the models. Since variance and correlation parts are separated, there are totally 4 types of hypotheses. They could be in general written as $H_0 : \theta_i = \theta_j$ versus $H_1 : \theta_i \neq \theta_j$. Below I describe all 4 types of hypotheses in detail. I test the sum of α and β (γ and δ), because one can interpret this sum as shock persistence.

(1) $\theta = \omega_1,\ \omega_2,\ \alpha_1,\ \alpha_2,\ \alpha_1+\beta_1,\ \alpha_2+\beta_2;\ i,j = MixCCC,\ MixDCC,\ MixDECO.$

(2) For "MixCCC", "MixDCC" and "MixDECO": $\theta = \omega,\ \alpha,\ \alpha+\beta;\ i=1,\ j=2.$

(3) $\theta = \lambda, \gamma_1,\ \gamma_2,\ \gamma_1+\delta_1,\ \gamma_2+\delta_2;\ i,j = MixCCC,\ MixDCC,\ MixDECO.$[6]

(4) For "MixDCC" and "MixDECO": $\theta = \gamma,\ \gamma+\delta;\ i=1,\ j=2.$

Table 4: Statistics of the estimates of the component probability and the correlation process parameters across the models

	λ_{CCC}	λ_{DCC}	γ_1	$\gamma_1+\delta_1$	γ_2	$\gamma_2+\delta_2$	λ_{DECO}	γ_1	$\gamma_1+\delta_1$	γ_2	$\gamma_2+\delta_2$
						Data Set 1					
Mean	0.8092	0.8224	0.0140	0.7119	0.0373	0.7054	0.7621	0.1711	0.8928	0.1067	0.8927
S.D.	0.0995	0.0918	0.0116	0.2654	0.0656	0.2968	0.1205	0.2400	0.1802	0.1761	0.1321
						Data Set 2					
Mean	0.7636	0.7666	0.0124	0.7606	0.0265	0.7746	0.7432	0.1572	0.8909	0.1246	0.8921
S.D.	0.1186	0.1150	0.0164	0.2265	0.0422	0.2435	0.1332	0.1889	0.2088	0.1728	0.1629
						Data Set 3					
Mean	0.7966	0.8035	0.0108	0.7640	0.0402	0.6810	0.7972	0.1088	0.8935	0.0666	0.8797
S.D.	0.1056	0.1021	0.0135	0.2767	0.0561	0.2857	0.1134	0.1734	0.1759	0.1261	0.1461
						Data Set 4					
Mean	0.7529	0.7558	0.0063	0.6878	0.0191	0.5659	0.7483	0.2126	0.9141	0.0450	0.8790
S.D.	0.1021	0.1039	0.0051	0.2474	0.0213	0.3573	0.1035	0.2307	0.1082	0.1164	0.1271
						Data Set 5					
Mean	0.7878	0.7945	0.0056	0.4700	0.0113	0.5365	0.8004	0.3122	0.8672	0.2143	0.8888
S.D.	0.0917	0.0946	0.0046	0.3541	0.0131	0.3901	0.0926	0.3061	0.2071	0.2923	0.1762

Note: Summary of the estimates of the component probability and of the correlation process parameters is presented for three types of the correlation matrix dynamics and for all five data samples. The mean and the standard deviation of parameters are calculated on the base of model estimates on 4795 rolling windows.

The asymptotic distribution of the difference in parameters is not known, therefore I approximate the distribution of the test statistic using block bootstrap sampling. The test statistic Y which I employ is a t-statistic of the following form:

$$Y = \frac{\sqrt{T}\hat{\mathbb{E}}[\theta_i - \theta_j]}{\sqrt{\hat{\mathbb{V}}_{HAC}[\theta_i - \theta_j]}}, \tag{35}$$

where $\hat{\mathbb{V}}_{HAC}$ is a heteroskedasticity and autocorrelation-consistent variance estimator, θ, i and j are defined above for the tests (1) to (4). The number of draws from the time series of the difference $(\theta_i - \theta_j)_t$ between parameter estimates is 1000, the block size is 5 observations (presuming domination of weekly patterns). If a certain hypothesis is not rejected on the 5% significance level, I indicate it by 0; if rejected, I indicate it by the sign of $\hat{\mathbb{E}}[\theta_i - \theta_j]$.

[6]For "MixCCC" in (3) θ takes value of λ only.

The results of the comparison of the GARCH parameters across the models (the Type 1 hypotheses) are presented in Table 5 (for all 5 stocks in Data Set 1). Unfortunately, there is no case, when the results of testing for a particular parameter are the same for all 5 stocks. But we can draw the following conclusions from the columns where a particular value prevails and no sign disagreement is present. (1) The variances of the first component of the mixture CCC-GARCH(1,1) react more to shocks in comparison to the mixture DCC(1,1)-GARCH(1,1), and for the variances of the second component the opposite is true. Additionally, the shocks in the second component are less persistent. (2) Comparing the mixture CCC-GARCH(1,1) to the mixture DECO(1,1)-GARCH(1,1), we see that the variances of the second component of the former model react more to shocks and the unconditional variance is higher. (3) Comparing the mixture DCC(1,1)-GARCH(1,1) and the mixture DECO(1,1)-GARCH(1,1), we see that the variances of the second component of the former model react more to shocks and the shocks in the second component are more persistent. To conclude, the noticeable difference in the estimates of the variance dynamics lies in the second component: for the mixture DCC(1,1)-GARCH(1,1) there is the highest reaction to shocks and the highest persistence, the next in this ranking is the mixture CCC-GARCH(1,1) and the last is the mixture DECO(1,1)-GARCH(1,1).

Table 5: Significance of differences in the GARCH parameter estimates across the models

	ω_1	α_1	$\alpha_1+\beta_1$	ω_2	α_2	$\alpha_2+\beta_2$
MixCCC vs. MixDCC						
# 1	-1	0	0	0	-1	0
# 2	0	1	0	1	0	-1
# 3	1	1	-1	0	-1	0
# 4	0	0	0	1	-1	-1
# 5	0	1	0	0	0	-1
MixCCC vs. MixDECO						
# 1	0	-1	0	1	1	1
# 2	0	1	0	0	0	0
# 3	-1	0	1	1	1	0
# 4	0	0	1	0	1	1
# 5	0	0	0	1	1	0
MixDCC vs. MixDECO						
# 1	1	-1	-1	1	1	1
# 2	0	0	0	0	0	1
# 3	-1	-1	1	1	1	0
# 4	0	0	1	-1	1	1
# 5	0	0	0	1	0	0

Note: The results of testing the hypotheses of type 1, $H_0: \theta_i = \theta_j$ versus $H_1: \theta_i \neq \theta_j$, $\theta = \omega_1,\ \omega_2,\ \alpha_1,\ \alpha_2,\ \alpha_1+\beta_1,\ \alpha_2+\beta_2$; $i,j = MixCCC, MixDCC, MixDECO$. If a certain hypothesis is not rejected on the 5% significance level, it is indicated by 0; if rejected, it is indicated by the sign of $\hat{\mathbb{E}}[\theta_i - \theta_j]$.

Table 6 shows results of testing the Type 2 hypotheses, namely, the cross-component parameter comparison. In this case the test results are much more coherent and show that the second component (with the probability less than 0.5) is in general more volatile, its variances react more to shocks and the shocks are more persistent. However, the last feature is less pronounced.

Table 6: Significance of differences in the GARCH parameter estimates across components

	MixCCC			MixDCC			MixDECO		
	ω	α	$\alpha+\beta$	ω	α	$\alpha+\beta$	ω	α	$\alpha+\beta$
# 1	-1	-1	-1	-1	-1	-1	-1	-1	-1
# 2	-1	-1	0	-1	-1	-1	-1	-1	0
# 3	-1	-1	1	-1	-1	1	-1	-1	1
# 4	-1	-1	-1	-1	-1	-1	-1	-1	-1
# 5	-1	-1	-1	-1	-1	-1	-1	-1	-1

Note: The results of testing the hypotheses of type 2, $H_0 : \theta_i = \theta_j$ versus $H_1 : \theta_i \neq \theta_j$; for MixCCC, MixDCC and MixDECO: $\theta = \omega,\ \alpha,\ \alpha+\beta$; $i = 1,\ j = 2$. If a certain hypothesis is not rejected on the 5% significance level, it is indicated by 0; if rejected, it is indicated by the sign of $\hat{\mathbb{E}}[\theta_i - \theta_j]$.

Table 7 presents the results of testing the Type 3 hypotheses for equality of the correlation process parameters and the component probabilities between specifications. One can conclude that the probability of the first component estimated in the CCC- and DECO-based specifications is on average smaller than in the DCC-based specification. There is no clear relation between the values of λ in the CCC- and DECO-based specifications.

One can notice a big contrast in the behaviour of the DCC and DECO estimates. The reaction to shocks is smaller for the DCC-based specification in both components, which is statistically confirmed (see Table 7, part MixDCC vs. MixDECO, columns γ_1 and γ_2). The same holds for the shock persistence (columns $\gamma_1 + \delta_1$ and $\gamma_2 + \delta_2$). The interesting thing is how the estimates change as the number of stocks in data sets increases (see Table 4, Data Set 2 to 4, Data Set 4 is composed of Data Set 2 and 3). The persistence of shocks in the DECO-based specification remains approximately the same and the reaction to shocks grows. For the DCC-based specification both the persistence and shock-reaction parameters decrease with the number of stocks in the data set. Both types of dependences on the number of stocks are actually not desirable.

Table 8 shows the results of cross-component comparison of parameter values for the DCC- and DECO-based specifications (the Type 4 hypotheses). It allows for concluding that the reaction to shocks in the second component is larger for the DCC-based specification, but smaller for the DECO-based specification. There is no clear conclusion for the persistence, since there is no sign agreement in the results of tests.

I tried to pursue tests for in-sample fitting, in particular, whether the mixture models fit better than their univariate prototypes. The tests were designed in the same way as described in the previous chapter. However I met the following difficulties: at the step where a mixture model must be fitted to the data simulated from a one-component based model, the EM algorithm often does not provide valid estimates for the second component. Therefore, this way to test these hypotheses is cumbersome and not fully appropriate. At the moment the literature still does not provide any other type of tests which is feasible for my family of models.

Table 7: Significances of difference in the DCC and λ parameter estimates across the models

	MixCCC vs. MixDCC	MixCCC vs. MixDECO	MixDCC vs. MixDECO				
	λ	λ	λ	γ_1	$\gamma_1+\delta_1$	γ_2	$\gamma_2+\delta_2$
Data Set 1	-1	1	1	-1	-1	-1	-1
Data Set 2	0	1	1	-1	-1	-1	-1
Data Set 3	-1	0	0	-1	-1	-1	-1
Data Set 4	-1	0	1	-1	-1	-1	-1
Data Set 5	-1	-1	0	-1	-1	-1	-1

Note: The results of testing the hypotheses of type 3, $H_0 : \theta_i = \theta_j$ versus $H_1 : \theta_i \neq \theta_j$, $\theta = \lambda, \gamma_1,\ \gamma_2,\ \gamma_1 + \delta_1,\ \gamma_2 + \delta_2$; $i, j = MixCCC,\ MixDCC, MixDECO$ (for MixCCC θ takes value of λ only). If a certain hypothesis is not rejected on the 5% significance level, it is indicated by 0; if rejected, it is indicated by the sign of $\hat{\mathbb{E}}[\theta_i - \theta_j]$.

Table 8: Significance of differences in the DCC parameter estimates across components

	MixDCC		MixDECO	
	γ	$\gamma+\delta$	γ	$\gamma+\delta$
Data Set 1	-1	0	1	0
Data Set 2	-1	0	1	0
Data Set 3	-1	0	1	0
Data Set 4	-1	1	1	1
Data Set 5	-1	-1	1	-1

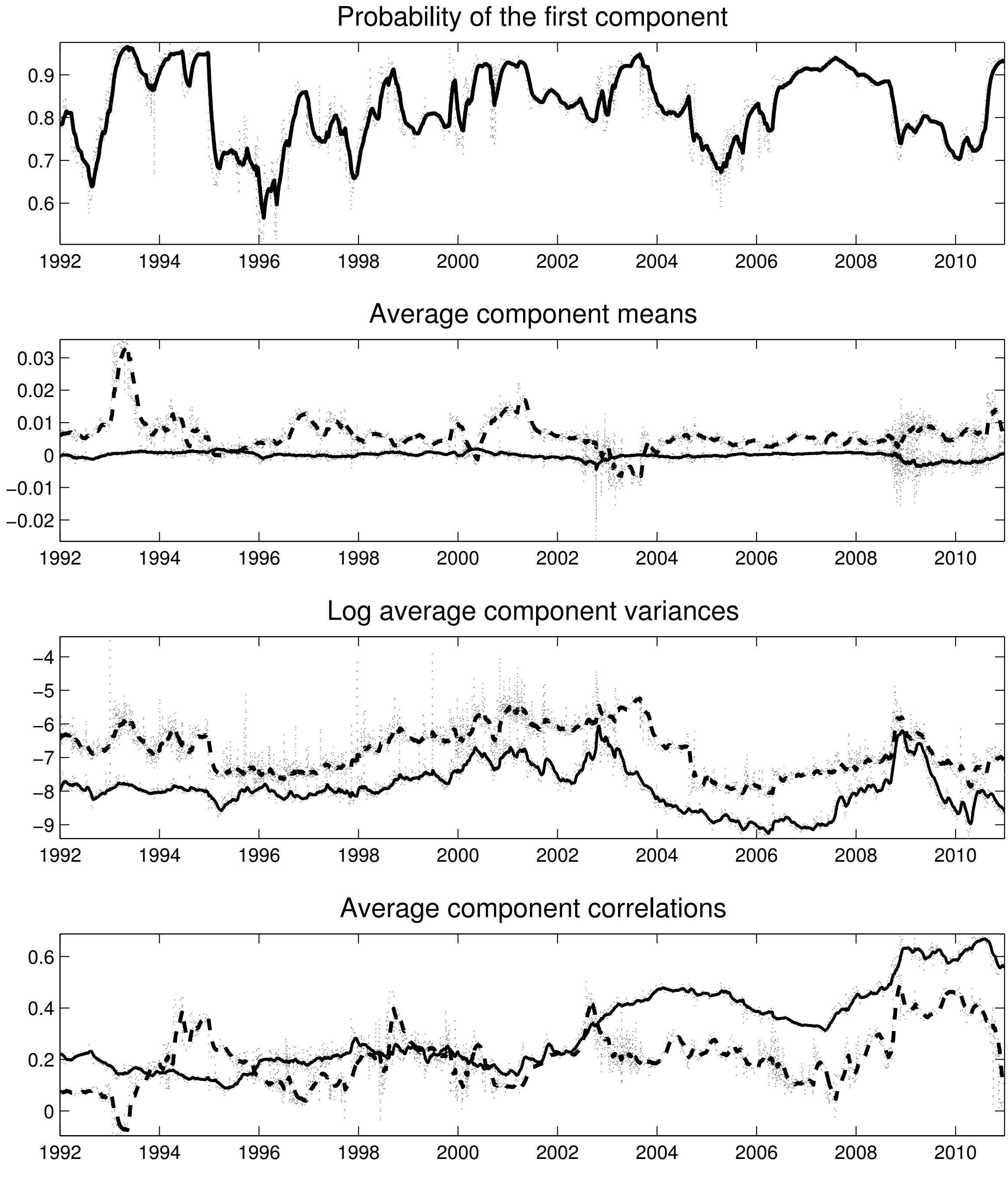
Note: The results of testing the hypotheses of type 4, $H_0 : \theta_i = \theta_j$ versus $H_1 : \theta_i \neq \theta_j$; for MixDCC and MixDECO: $\theta = \lambda,\ \gamma,\ \gamma + \delta$; $i = 1,\ j = 2$. If a certain hypothesis is not rejected on the 5% significance level, it is indicated by 0; if rejected, it is indicated by the sign of $\hat{\mathbb{E}}[\theta_i - \theta_j]$.

Next I illustrate the dynamic behaviour of the two components using the mixture DCC(1,1)-GARCH(1,1) model (the results for other two models are very similar). Each figure corresponding to the data sets from 1 to 3 consists of four graphs. The graphs show the time evolution of different estimates on the rolling window. The thin lines in the background correspond to the factual estimates, which are often fluctuating, whereas the thick lines show exponentially smoothed estimates in order to emphasize the overall trend. The solid line always corresponds to the first component (with a greater probability) and the dashed line corresponds to the second component. The first graph from the top presents the estimates of the first component probability. The second graph shows the estimates of the component means averaged across stocks. The third graph demonstrates the one-day ahead predictions of the component return variances averaged across stocks (log is taken afterwards to balance the heterogeneity), whereas the last graph presents averaged predicted correlations.

The three figures have the following similarities: (1) About two-three years before the recent financial crisis at the end of 2008 the probability of the first component was about 0.9; (2) The average mean of the second component is most of the time greater than the average mean of the first component and its values are more volatile; (3) The average predicted variance of the second component is almost always higher that the average predicted variance of the first component; (4) The behaviour of the average predicted correlations changes dramatically after 2002. Before this date the average correlations of two components were close to each other and none of them always remained higher than the other. After this date the distance between the average predicted correlations begins to grow rapidly and the average correlation of the first component is higher than that of the second component.

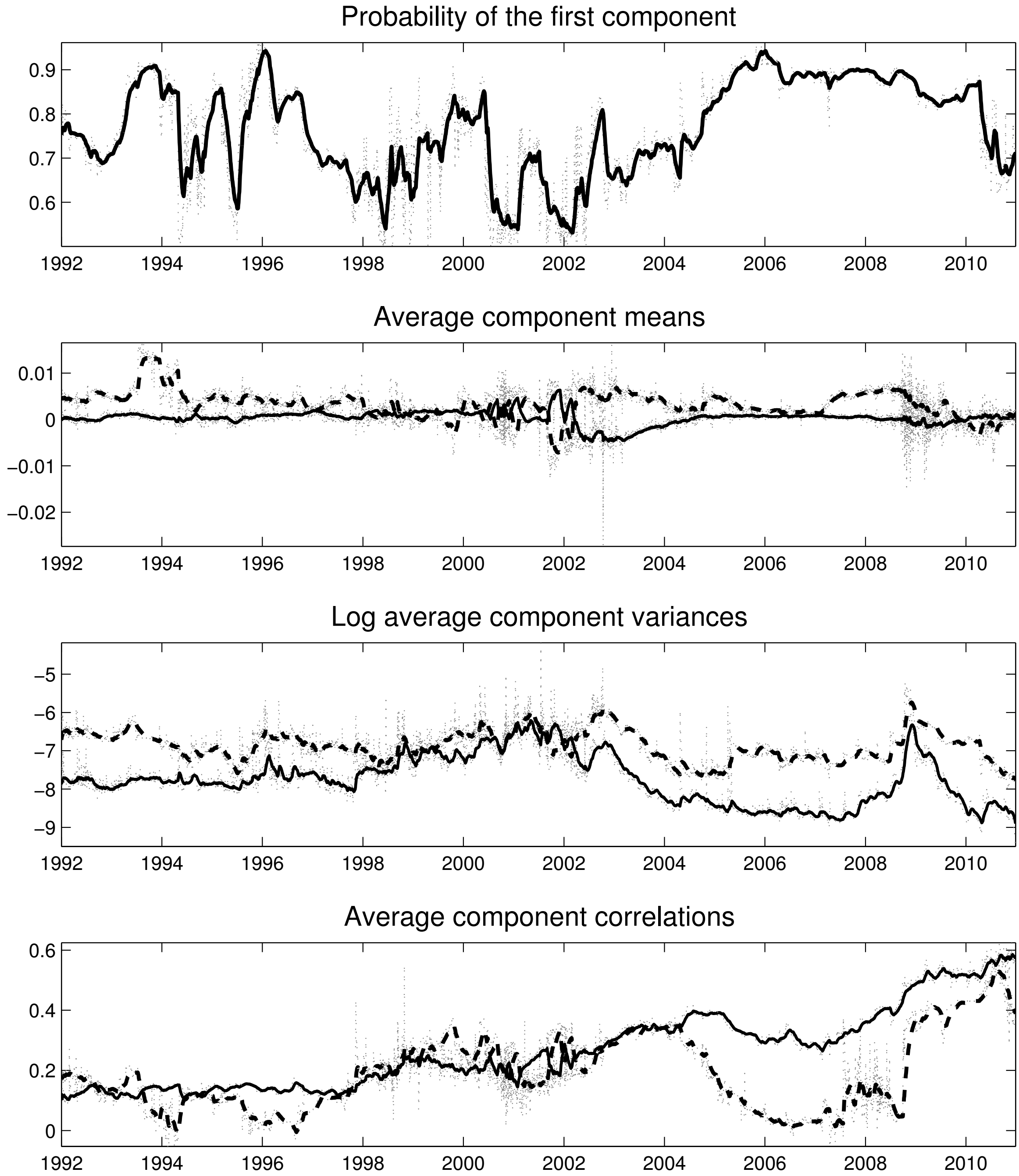
On the first glance, these findings seem to be in contradiction with arguments of Longin and Solnik, 2001 and Ang and Chen, 2002, who conclude that the market is likely to have two regimes: (1) positive returns, low volatilities and correlations and (2) negative returns, high volatilities and correlations. We should be careful with the interpretation of the results in this article. First, we should not mix regime-switching models and finite mixture models, although in theory the latter class is a subset of the former class. Often in the literature the dynamics of the market is interpreted as the "momentum" regime-switching process which means that transition to the other state is less likely than remaining in the same state. The finite mixture models mix the components unconditionally on the state where the market is right now. Therefore, the behaviour of the finite mixture components does not have to be the same as the behaviour of the regime-switching states.

Figure 2: Data Set 1, the mixture DCC(1,1)-GARCH(1,1)



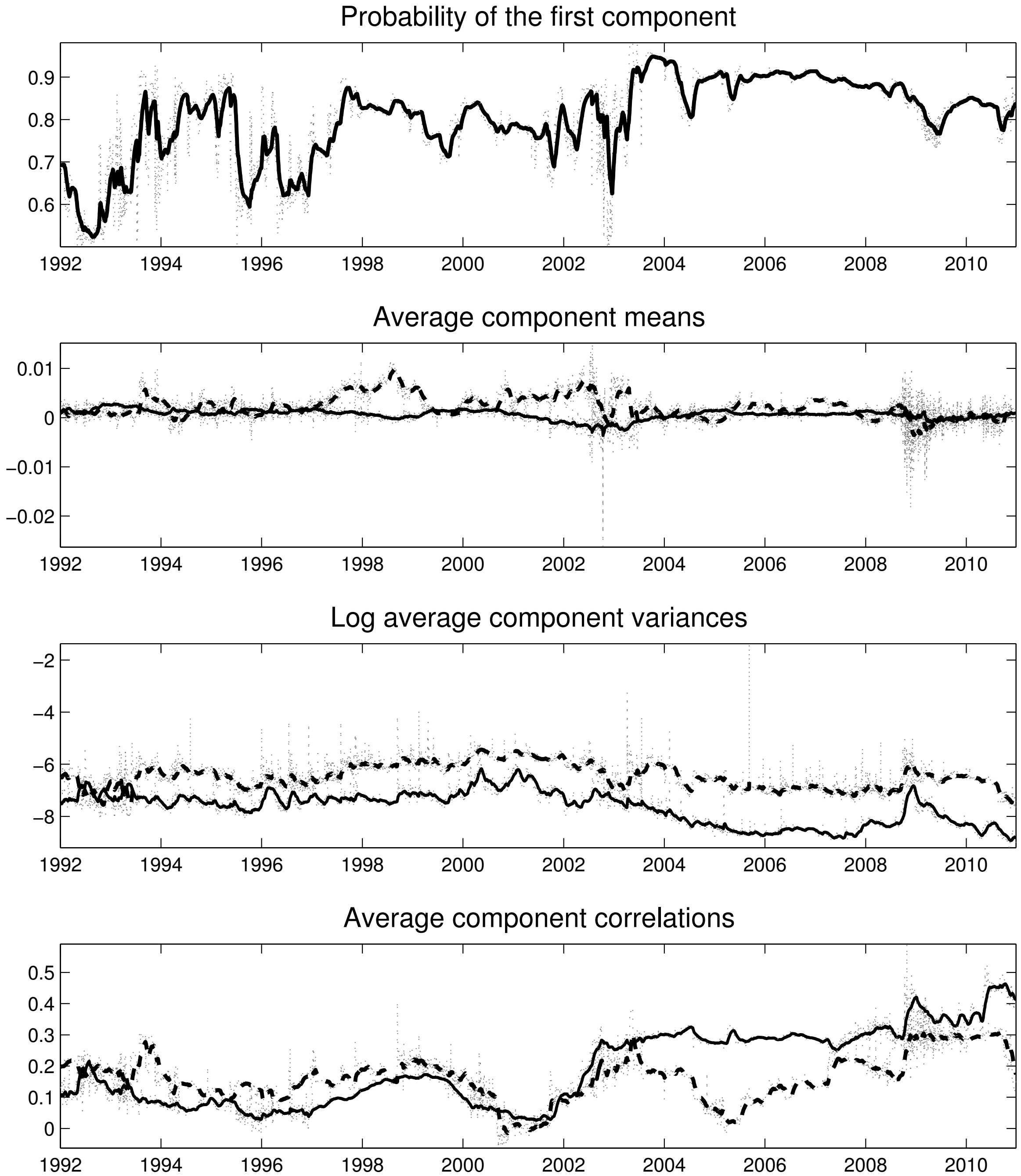
Note: The solid line refers to the 1st component, the dashed line refers to the 2d component. From the top, Graph 1: The first component probability; Graph 2: The component means averaged across stocks; Graph 3: Log of the one-day ahead predictions of the component return variances averaged across stocks; Graph 4: Stock-averaged one-day ahead predicted correlations.

Figure 3: Data Set 2, the mixture DCC(1,1)-GARCH(1,1)



Note: The solid line refers to the 1st component, the dashed line refers to the 2d component. From the top, Graph 1: The first component probability; Graph 2: The component means averaged across stocks; Graph 3: Log of the one-day ahead predictions of the component return variances averaged across stocks; Graph 4: Stock-averaged one-day ahead predicted correlations.

Figure 4: Data Set 3, the mixture DCC(1,1)-GARCH(1,1)



Note: The solid line refers to the 1st component, the dashed line refers to the 2d component. From the top, Graph 1: The first component probability; Graph 2: The component means averaged across stocks; Graph 3: Log of the one-day ahead predictions of the component return variances averaged across stocks; Graph 4: Stock-averaged one-day ahead predicted correlations.

If we analyze the dynamics of the average unconditional variance and correlation during the crises of 1998, 2002 and 2008 years, we see that there is no contradiction with the regime-switching model, since the variances and the correlations of both components rise. In general we see that the first component has the average mean close to zero, a low average variance, but a high average correlation. The second component has a positive average mean, a high averaged variance, and a low average correlation. This structure become very pronounced after 2002, because the tendency of stock prices to move together increased with the development of electronic trading systems and telecommunication technologies. A possible interpretation is that the first component bears the correlation risk and the second component bears the variance risk, and these two types of risk are mixed.

It is worth mentioning that we interpret the average values for each component. In general, the component estimates are very heterogeneous: the mean vectors contain positive as well as negative values within the same component. Some variances and correlations could be lower than those of the other component and some could be higher at the same time.

Finally, I address the issue of the predictive power of the suggested family of models. To evaluate the quality of the density forecasts, I apply the logarithmic scoring rule approach (some good examples of its application could be found in Mitchell and Hall, 2005). If $\hat{f}_{t+1}(x)$ is a one-step-ahead stock return density forecast based on the information set Ω_t, then the logarithmic scoring rule is

$$S(\hat{f}_{t+1}(x); \boldsymbol{r}_{t+1}) = \log \hat{f}_{t+1}(\boldsymbol{r}_{t+1}). \tag{36}$$

Density forecasts coming from two different models i and j could be then compared by testing whether the difference of their scoring rules is statistically significant. Denote by $\bar{d}_{i,j,t+1} = \hat{\mathbb{E}}[S(\hat{f}_{i,t+1}(x); \boldsymbol{r}_{t+1}) - S(\hat{f}_{j,t+1}(x); \boldsymbol{r}_{t+1})]$ the sample average of scoring rule differences. Then, under the null hypothesis $\mathbb{E}[S(\hat{f}_{i,t+1}(x); \boldsymbol{r}_{t+1}) - S(\hat{f}_{j,t+1}(x); \boldsymbol{r}_{t+1})] = 0$, the following statistic asymptotically follows the standard normal distribution under some proper assumptions (see Giacomini and White, 2006 for the proof):

$$Y_U = \frac{\sqrt{T}\bar{d}_{i,j,t+1}}{\sqrt{\hat{\mathbb{V}}_{HAC}[\sqrt{T}\bar{d}_{i,j,t+1}]}}, \tag{37}$$

where $\hat{\mathbb{V}}_{HAC}$ is a heteroskedasticity and autocorrelation-consistent variance estimator.

Giacomini and White, 2006 distinguish between unconditional and conditional approaches to predictive ability testing. The statistic described above refers to the unconditional approach.

The authors point out that a model that well approximates the data-generating process may forecast poorly. It might make sense to consider the quality of predictions conditional on the information set available at time t. To construct the appropriate statistic, one should choose a set of functions h_t, which are measurable with respect to the information set. In this paper we take the most simple one: $h_t = \bar{d}_{i,j,t}$. If we denote by $Z_{i,j,t}$ the vector $[\bar{d}_{i,j,t+1} \; \bar{d}_{i,j,t}]$, then the corresponding Wald-type statistic is

$$Y_C = TZ_{i,j,t}\hat{\mathbb{V}}_{HAC}[\sqrt{T}Z_{i,j,t}]^{-1}Z'_{i,j,t}, \tag{38}$$

which under the null hypothesis follows the chi-squared distribution with 1 degree of freedom (in the general case q degrees of freedom, where q is the number of measurable functions in the set).

Additionally, to compare the predictive ability in the left tail, which is very important feature for financial applications, I employ the approach of Diks et al., 2011, who suggested the use of properly specified weighted logarithmic scoring rules. The weighted scoring rule applied in this paper has the formula

$$S^{csl}(\hat{f}_t; \boldsymbol{\varepsilon}_{t+1}) = 1(\boldsymbol{\varepsilon}_{t+1} \in A_t) \log \hat{f}_t(\boldsymbol{\varepsilon}_{t+1}; \boldsymbol{\Psi}) + 1(\boldsymbol{\varepsilon}_{t+1} \in \bar{A}_t) \log \left(\int_{\bar{A}_t} \hat{f}_t(\boldsymbol{\varepsilon}_{t+1}; \boldsymbol{\Psi}) \right), \tag{39}$$

where A_t denotes a set, which definition is model-independent, and $\bar{A}_t$ is its complement.

This scoring rule considers the accuracy of the density forecast for the total probability of the region of interest by means similar to these of the tobit regression (Tobin, 1958), because we do not care about the exact shape of the distribution outside of the tail, but still consider that the overall probability of being in this region matches the reality. Diks et al., 2011 found that this scoring rule also leads to more powerful tests, because more information is used.

There are numerous ways to cut the left tail for multivariate distributions. I define an approximate dynamic 5% left tail A_t as follows: I look for such a constant $\bar{\alpha}$, so that

$$\frac{1}{T-l} \sum_{t=l+1}^{T} \prod_{i=1}^{n} 1[\boldsymbol{r}_{t,i} \leq \hat{Q}(\bar{\alpha})_{t,i}] \approx 0.05, \tag{40}$$

where l is the window length, $\hat{Q}(\alpha)_{t,i}$ is the value of the α percent quantile for r_i calculated on the window $[t-l, t-1]$. This definition is the empirical counterpart of $\mathbb{E}[1(\boldsymbol{\varepsilon}_{t+1} \in A_t)] = 0.05$. The values of $\bar{\alpha}$ are the following: 0.35 for Data Set 1, 0.37 for Data Set 2, 0.42 for Data Set 3, 0.57 for Data Set 4 and 0.63 for Data Set 5.

The choice of the competing models is determined by two factors: feasibility of parameter estimates on the considered data sets and the overall attention and popularity in the literature. It resulted in 5 competing models to predict the distribution of the stock returns. The first one is simply a multivariate normal distribution without any dynamics (denoted "Normal"). The second one is a two-component mixture of multivariate normal distributions (denoted "MixNorm"). In these two cases, the density estimated on the rolling window is used for the one-step ahead prediction. Another three models are CCC, DCC and DECO respectively, in their classical formulations. I estimate their parameters on the rolling window and do one-step ahead prediction of the parameters of the multivariate normal distribution to obtain the prediction of the stock return density. The density prediction for the class of models introduced in this paper is pursued in the same way.

Table 9: Comparing predictive ability: MixNorm versus Normal and the CC class

#	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
Unconditional predictive ability, simple scoring rule					
Normal	6.76 (0.0000)	7.15 (0.0000)	7.72 (0.0000)	7.61 (0.0000)	7.92 (0.0000)
CCC	-1.31 (0.1875)	1.36 (0.1744)	1.54 (0.1231)	0.80 (0.4247)	0.99 (0.3222)
DCC	-1.46 (0.1418)	1.32 (0.1880)	1.38 (0.1672)	0.65 (0.5152)	0.84 (0.3985)
DECO	0.49 (0.6175)	2.90 (0.0037)	1.83 (0.0669)	3.25 (0.0011)	2.01 (0.0447)
Conditional predictive ability, simple scoring rule					
Normal	46.09 (0.0000)	68.31 (0.0000)	79.49 (0.0000)	84.8 (0.0000)	87.81 (0.0000)
CCC	3.11 (0.0777)	2.48 (0.1151)	5.32 (0.0211)	3.07 (0.0798)	6.66 (0.0099)
DCC	3.78 (0.0517)	2.16 (0.1415)	4.68 (0.0305)	2.80 (0.0945)	6.55 (0.0105)
DECO	3.29 (0.0694)	9.48 (0.0021)	6.41 (0.0114)	15.40 (0.0001)	15.49 (0.0001)
Unconditional predictive ability, weighted scoring rule					
Normal	2.46 (0.0140)	2.01 (0.0443)	3.13 (0.0018)	3.14 (0.0017)	4.50 (0.0000)
CCC	2.07 (0.0382)	1.03 (0.3014)	1.92 (0.0555)	4.11 (0.0000)	3.56 (0.0004)
DCC	1.99 (0.0458)	1.02 (0.3063)	1.73 (0.0837)	3.88 (0.0001)	3.39 (0.0007)
DECO	2.67 (0.0077)	0.90 (0.3657)	1.98 (0.0473)	4.59 (0.0000)	4.24 (0.0000)

Note: The table presents the results of comparing the predictive ability of the static two-component mixture of normal distributions to that of the other four competing models. For each pair and for each of the three testing methods the value of the test statistic and the corresponding p-value (in parentheses) are given.

The first interesting result that we obtain comparing scoring rules for 5 competing models is presented in Table 9. It turns out that there is no clear hierarchy across these models, except that the model "Normal" is pronouncedly dominated according to all three criteria. The model "MixNorm" leads in the left tail predictions for all the data sets except Data Set 2, but as for the full density prediction, conditionally and unconditionally evaluated, the null hypothesis is not rejected for the most pairs in the table. Since the suggested class of models amalgamates distinct features of the finite mixture distribution family and of the conditional correlation family, these results provide additional optimism about the performance of the mixture normal conditional correlation models.

Table 10 presents statistical properties of the scoring rule values for the eight considered models and all the five data sets. The properties include the mean, the standard deviation, the minimum and maximum values. We observe the following from the table. (1) The "MixDECO" has the highest minimum values for the data sets consisting of 10 and 15 stocks. It means that this model makes less severe prediction errors. (2) The "MixDECO" also has the lowest standard deviations for three of five data sets, which could be interpreted as more stable performance. (3) The "MixCCC" has the highest maximum values for three data sets of five. It means that this model is able to make the most precise predictions in some cases. (4) The "MixDCC" has the highest average value of the scoring rules for four data sets of five, therefore, it makes on average the most precise predictions.

Table 10: Statistical properties of simple scoring rules

	Normal	MixNorm	CCC	DCC	DECO	MixCCC	MixDCC	MixDECO
				Data Set 1				
Mean	-8.8695	-8.5130	-8.4692	-8.4639	-8.5292	-8.3073	-8.3015	-8.3845
S.D.	5.1823	3.5918	4.2320	4.2479	4.2025	3.2950	3.3145	3.3186
Max	-3.7642	-2.9806	-2.8950	-2.9603	-2.9188	-1.8159	-2.1139	-2.3419
Min	-200.6186	-101.8252	-148.2679	-147.4740	-148.3671	-64.2042	-67.6009	-66.2850
				Data Set 2				
Mean	-9.1681	-8.9050	-8.9558	-8.9547	-9.0055	-8.8165	-8.8058	-8.8868
S.D.	3.7686	2.9984	3.9943	4.0191	3.8361	3.2613	3.1671	3.1490
Max	-4.6152	-3.9515	-3.4686	-3.4661	-3.4883	-1.8688	-2.2200	-2.8815
Min	-76.0576	-38.7737	-133.6215	-133.9595	-108.0928	-87.5990	-67.9598	-79.6961
				Data Set 3				
Mean	-9.2652	-8.9338	-8.9802	-8.9756	-8.9880	-8.8007	-8.7915	-8.8166
S.D.	4.3664	3.2336	3.8580	3.8619	3.8107	3.2384	3.2474	3.1994
Max	-4.1211	-3.3445	-3.4174	-3.4224	-3.3989	-2.6769	-2.6896	-2.6098
Min	-139.8898	-67.6736	-86.6243	-87.8248	-82.7391	-63.9706	-66.0316	-72.9673
				Data Set 4				
Mean	-18.1418	-17.6240	-17.6642	-17.6570	-17.7782	-17.4219	-17.4113	-17.5086
S.D.	6.6253	5.4835	6.2968	6.3088	6.1496	5.5766	5.4969	5.2791
Max	-8.8727	-7.2538	-6.4473	-6.4543	-6.6574	-4.8030	-4.7604	-5.5191
Min	-156.8873	-94.3834	-148.3348	-151.6454	-116.7992	-96.8460	-84.1368	-84.0622
				Data Set 5				
Mean	-24.9395	-24.1731	-24.2450	-24.2345	-24.3264	-23.9463	-23.9392	-23.9179
S.D.	9.6096	7.8121	9.4565	9.4422	9.3428	8.1892	8.2174	7.1964
Max	-11.7319	-9.9906	-8.5641	-8.6221	-9.1996	-6.7096	-6.7361	-7.7595
Min	-174.7697	-122.5897	-219.1795	-219.8523	-228.6646	-162.7093	-154.6145	-108.5021

Note: For each of the eight models the mean, the standard deviation, the minimum and the maximum of the simple scoring rule (formula (36)) are reported. These statistics are calculated across the time dimension for all the five data sets.

Table 11 and Table 12 give the results of the tests for the unconditional and conditional predictive ability using the simple scoring rule. Since the "MixDCC" often has the highest average value of the scoring rule, I only test how this particular model performs compared to all other models instead of reporting the results for all possible pairs (for brevity). Consideration of p-values in Table 11 and Table 12 allows for concluding that, first, the class of multivariate normal conditional correlation models dominates the competitors in terms of conditional and unconditional predictability. Second, the predictive abilities of the "MixCCC" and the "MixDCC" are statistically indistinguishable for almost all data sets. Moreover, for Data Set 5 and the test for the unconditional predictive ability, the "MixDCC" and the "MixDECO" also perform equally well.

Table 11: Unconditional predictive ability test, simple scoring rule

#	Normal	MixNormal	CCC	DCC	DECO	MixCCC	MixDECO
1	7.75 (0.0000)	7.41 (0.0000)	6.10 (0.0000)	5.90 (0.0000)	7.80 (0.0000)	1.32 (0.1873)	5.50 (0.0000)
2	7.68 (0.0000)	4.79 (0.0000)	5.50 (0.0000)	5.36 (0.0000)	7.43 (0.0000)	1.86 (0.0624)	6.18 (0.0000)
3	8.56 (0.0000)	6.66 (0.0000)	7.55 (0.0000)	7.40 (0.0000)	7.89 (0.0000)	1.86 (0.0626)	2.71 (0.0068)
4	8.12 (0.0000)	6.16 (0.0000)	6.59 (0.0000)	6.38 (0.0000)	9.05 (0.0000)	0.85 (0.3981)	4.03 (0.0001)
5	8.07 (0.0000)	4.59 (0.0000)	6.24 (0.0000)	6.10 (0.0000)	6.98 (0.0000)	0.37 (0.7118)	-0.46 (0.6491)

Note: The table reports the statistic Y_U calculated using the simple scoring rule (36). The "MixDCC" model is contraposed to other seven competing models. The corresponding p-values are given in parentheses.

Table 12: Conditional predictability test of Giacomini and White (2006)

#	Normal	MixNormal	CCC	DCC	DECO	MixCCC	MixDECO
1	61.0 (0.0000)	51.4 (0.0000)	39.4 (0.0000)	36.8 (0.0000)	65.1 (0.0000)	2.1 (0.1498)	31.1 (0.0000)
2	87.8 (0.0000)	23.4 (0.0000)	30.3 (0.0000)	28.2 (0.0000)	55.8 (0.0000)	3.6 (0.0592)	39.2 (0.0000)
3	100.2 (0.0000)	46.4 (0.0000)	55.8 (0.0000)	54.0 (0.0000)	61.5 (0.0000)	4.2 (0.0396)	9.1 (0.0026)
4	99.6 (0.0000)	40.7 (0.0000)	40.2 (0.0000)	37.8 (0.0000)	80.5 (0.0000)	1.3 (0.2542)	24.6 (0.0000)
5	97.0 (0.0000)	20.5 (0.0000)	36.7 (0.0000)	35.3 (0.0000)	45.7 (0.0000)	2.2 (0.1352)	11.2 (0.0008)

Note: The table reports the statistic Y_C calculated using the simple scoring rule (36). The "MixDCC" model is contraposed to other seven competing models. The corresponding p-values are given in parentheses.

Table 13 presents the same four basic statistical properties of the weighted scoring rule values as Table 10 does. Now the model ranking with respect to the statistics differ in some aspects. (1) The "MixDCC" still has the highest mean value of the weighted scoring rule for four data sets of five (except Data Set 4). (2) The "MixDCC" also has the lowest standard deviation of the weighted scoring rule for three of five data sets (the "MixCCC" has the lowest one for the other two data sets). Since for the simple scoring rule the "MixDECO" is the leader, the current result means that the stability of the performance of this model is not preserved in the tail. (3) The "MixDECO" has the highest maximum values of the weighted scoring rule, meaning that it sometimes provides the highest value for the predicted probability of not being in the left tail. (4) Finally, the "MixCCC" has the highest minimum value of the weighted scoring rule in all data sets except of Data Set 3. Again, for the data sets consisting of 10 and 15 stocks the "MixDECO" has the highest minimum values of the simple scoring rule. It seems that the enforcement of equal correlations deteriorates the performance of the model in the tail.

Table 13: Statistical properties of weighted scoring rules

	Normal	MixNorm	CCC	DCC	DECO	MixCCC	MixDCC	MixDECO
				Data Set 1				
Mean	-0.5564	-0.5359	-0.5615	-0.5607	-0.5671	-0.5126	-0.5114	-0.5181
S.D.	2.4997	2.3257	2.5290	2.5275	2.5678	2.1913	2.1838	2.2333
Max	-0.0065	-0.0054	-0.0073	-0.0072	-0.0071	-0.0088	-0.0084	-0.0021
Min	-50.3898	-40.4665	-64.1934	-64.3903	-65.3838	-25.3625	-25.6753	-26.8056
				Data Set 2				
Mean	-0.5412	-0.5234	-0.5312	-0.5311	-0.5304	-0.5123	-0.5111	-0.5696
S.D.	2.4832	2.3311	2.3608	2.3580	2.3398	2.2829	2.2844	2.7423
Max	-0.0081	-0.0062	-0.0107	-0.0106	-0.0106	-0.0053	-0.0050	-0.0001
Min	-45.6668	-38.7737	-39.4304	-39.6264	-37.9561	-37.4633	-41.2888	-42.4466
				Data Set 3				
Mean	-0.5585	-0.5221	-0.5355	-0.5338	-0.5357	-0.5118	-0.5074	-0.5176
S.D.	2.6073	2.2344	2.2997	2.2755	2.2899	2.1858	2.1566	2.2369
Max	-0.0109	-0.0090	-0.0080	-0.0078	-0.0078	-0.0088	-0.0085	-0.0064
Min	-55.4561	-34.8857	-42.5833	-40.7070	-41.2148	-30.5334	-28.0246	-32.3836
				Data Set 4				
Mean	-0.8886	-0.8528	-0.8865	-0.8849	-0.8923	-0.8400	-0.8393	-0.8602
S.D.	4.0776	3.8213	3.9386	3.9245	3.9388	3.7688	3.7711	3.8941
Max	-0.0045	-0.0054	-0.0058	-0.0058	-0.0057	-0.0049	-0.0050	-0.0014
Min	-63.9189	-56.4940	-53.1484	-52.3822	-51.0857	-52.0084	-52.4441	-56.9473
				Data Set 5				
Mean	-1.2520	-1.1722	-1.2589	-1.2533	-1.2828	-1.1374	-1.1366	-1.1942
S.D.	5.7971	5.3244	5.8546	5.8285	5.9649	5.1520	5.1353	5.4008
Max	-0.0011	-0.0007	-0.0017	-0.0017	-0.0021	-0.0016	-0.0014	-0.0005
Min	-81.3725	-65.8866	-110.2407	-111.3220	-123.7107	-51.6499	-52.4042	-70.6199

Note: For each of the eight models the mean, the standard deviation, the minimum and the maximum of the weighted scoring rule (formula (39)) are reported. These statistics are calculated across the time dimension for all the five data sets.

Table 14 shows that the results of the unconditional predictive ability tests using the weighted scoring rule agree with the results of the same tests using the simple scoring rule. The "MixDCC" clearly dominates all the five competing models from the literature and also the "MixDECO" model. Still, despite on average the weighted score rule is the highest for the "MixDCC", the difference in the values of this rule between the "MixDCC" and "MixCCC" models is statistically negligible for four of five data sets.

Table 14: Unconditional predictive ability test, weighted scoring rule

#	Normal	MixNormal	CCC	DCC	DECO	MixCCC	MixDECO
1	3.50 (0.0005)	3.41 (0.0006)	4.56 (0.0000)	4.50 (0.0000)	4.74 (0.0000)	1.07 (0.2851)	2.03 (0.0421)
2	2.62 (0.0089)	2.36 (0.0185)	3.18 (0.0014)	3.15 (0.0016)	2.99 (0.0028)	0.71 (0.4750)	3.57 (0.0004)
3	3.50 (0.0005)	3.44 (0.0006)	3.79 (0.0001)	3.77 (0.0002)	3.90 (0.0001)	3.68 (0.0002)	2.93 (0.0033)
4	3.88 (0.0001)	2.97 (0.0030)	5.60 (0.0000)	5.46 (0.0000)	5.82 (0.0000)	0.52 (0.6045)	3.95 (0.0001)
5	4.98 (0.0000)	4.01 (0.0001)	4.69 (0.0000)	4.57 (0.0000)	5.22 (0.0000)	0.24 (0.8101)	5.20 (0.0000)

Note: The table reports the statistic Y_U calculated using the weighted scoring rule (39). The "MixDCC" model is contraposed to other seven competing models. The corresponding p-values are given in parentheses.

To summarize, the "MixCCC" and the "MixDCC" seem to provide the best predictions according to three types of tests for predictive ability. Despite the "MixDCC" is predicting slightly better, the "MixCCC" might be preferred because of its much faster estimation algorithm (about 8 times faster).

7 Conclusion

I suggest a class of mixture normal conditional correlation models which allows to exploit the advantages of finite mixtures and conditional correlation processes simultaneously. The finite mixtures are flexible, relatively easy to estimate and well-tractable if applied to some financial processes, such as stock returns. The model family presented in this paper allows to overcome the curse of dimensionality. Due to a convenient covariance matrix decomposition and a multi-step estimation algorithm these multivariate models could be estimated for a large-dimensional time series (given a sufficient number of observation available to identify the component parameters). The clearly described estimation algorithm is based on the EM-algorithm of Dempster et al., 1977. The algorithm has proven itself to be reliable during numerous numerical experiments.

The predictive ability of the three suggested models is compared to that of their five feasible competitors: a static multivariate normal distribution, a static finite mixture of multivariate normal distributions, the constant conditional correlation model of Bollerslev, 1990, the dynamic conditional correlation model of Engle, 2002 and the dynamic equicorrelation model of Engle and Kelly, 2012. Three different types of tests (including the test for the predictability in the left tail, important for financial applications) are performed on five data sets including up to 15 stock returns modeled simultaneously. The tests show that the mixture normal constant conditional correlation model and the mixture normal dynamic conditional correlation model confidently dominate all the competitors. The mixture normal dynamic equicorrelation model also predicts better than the competitors but its performance is worse than that of the two other models in the class. In general, the mixture normal dynamic conditional correlation model has the best out-of sample performance, but it is often statistically indistinguishable from the performance of the the mixture normal constant conditional correlation model. Additionally, the estimation of the latter is about eight times faster.

The most important obstacle for the research in this area is the absence of a proof of consistency of parameter estimates for the dynamic conditional correlation models. Aielli, 2011 provides only the heuristic proof for the corrected dynamic conditional correlation model of Engle, 2002. Another important research question is to investigate whether the proposed class of models provide some gains for risk management and portfolio selection. Finally, it would be interesting to explore the situation where the number of dimensions of a time series is comparable to the number of observations and find a way to tailor the estimation algorithm to work in this case of nearly singular covariance matrices of the mixture components.

References

[1] G. P. Aielli. Dynamic conditional correlation: On properties and estimation. July 2011.

[2] C. Alexander and E. Lazar. Normal mixture GARCH(1,1): applications to exchange rate modeling. *Journal of Applied Econometrics*, 21:307–336, 2006.

[3] A. Ang and J. Chen. Asymmetric correlation of equity portfolios. *Journal of Financial Economics*, 63:443–494, 2002.

[4] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Thinking coherently. *RISK*, 10:68–71, November 1997.

[5] L. Bauwens, C.M. Hafner, and J.V.K. Rombouts. Multivariate mixed normal conditional heteroskedasticity. *Computational Statistics & Data Analysis*, 51:3551–3566, 2007.

[6] L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

[7] V. S. Bawa and E. B. Lindenberg. Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics*, 5(2):189–200, 1977.

[8] D. Bertsimas, G. Lauprete, and A. Samarov. Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics & Control*, 28:1353–1381, 2004.

[9] A. Biglova, S. Ortobelli, S. T. Rachev, and S. Stoyanov. Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management*, 31(1):103–112, 2004.

[10] T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics*, 72:498–505, 1990.

[11] T. Bollerslev, R. Chou, and K. Kroner. ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1–2):5–59, 1992.

[12] T. Bollerslev, R. Engle, and J. M. Woolridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

[13] T. Bollerslev, U. Kretschmer, C. Pigorsch, and G. Tauchen. A discrete-time model for daily S & P 500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, 150(2):151–166, 2009.

[14] S. Brown and J. Warner. Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(1):3–31, 1985.

[15] T. F. Crack and O. Ledoit. Using central limit theorems for dependent data. *Journal of Financial Education*, 36:38–60, 2010.

[16] S. Crawford. An application of the Laplace method to finite mixture distributions. *Journal of American Statistical Association*, 89:259–267, 1994.

[17] A.P. Dempster, N.M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society B*, 39:1–38, 1977.

[18] C. Diks, V. Panchenko, and D. van Dijk. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, 163:215–230, 2011.

[19] R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedastisity models. *Journal of Business and Economic Statistics*, 20(3):339–350, 2002.

[20] R. Engle. *Anticipating correlations*. Princeton University Press, 2009.

[21] R. Engle and B. Kelly. Dynamic equicorrelation. *Journal of Business and Economic Statistics*, 30(2):212–228, 2012.

[22] R. Engle and K. Kroner. Multivariate simultaneous GARCH. *Econometric Theory*, 11:122–150, 1995.

[23] R. Engle and K. Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. NBER Working Paper 8554, 2001.

[24] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

[25] E. F. Fama. Efficient capital markets: Ii. *Journal of Finance*, 46:1575–1617, 1991.

[26] C. Francq and J.-M. Zakoïan. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons Ltd, 2010.

[27] P. Galeano and M.C. Ausin. The gaussian mixture dynamic conditional correlation model: Parameter estimation, value at risk calculation, and portfolio selection. *Journal of Business & Economic Statistics*, 28(4):559–571, 2010.

[28] R. Giacomini and H. White. Tests for conditional predictive ability. *Econometrica*, 74:1545–1578, 2006.

[29] A. Y. Gu and J. Finnerty. The evolution of market efficiency: 103 years daily data of the dow. *Review of Quantitative Finance and Accounting*, 18:219–237, 2002.

[30] M. Haas, S. Mittnik, and M. S. Paolella. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, 2:493–530, 2004.

[31] M. Haas, S. Mittnik, and M. S. Paolella. Asymmetric multivariate normal mixture GARCH. *Comput. Stat. Data Anal.*, 53(6):2129–2154, 2009.

[32] W. V. Harlow and K.S. Rao. Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence. *Journal of Financial and Quantitative Analysis*, 24(3):285–311, 1989.

[33] E. Jacquier, N. G. Polson, and Rossi P.E. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 20(1):69–87, 2002.

[34] R. Jarrow and F. Zhao. Downside loss aversion and portfolio management. *Management Science*, 52(4):558–566, 2006.

[35] S. Kon. Models of stock returns – a comparison. *Journal of Finance*, 39(1):147–165, March 1984.

[36] F. Longin and B. Solnik. Extreme correlation of international equity markets. *Journal of Finance*, 56:649–676, 2001.

[37] H. M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[38] G. McLachlan and D. Peel. *Finite Mixture Models.* John Wiley and Sons, Inc., 2000.

[39] G. J. McLachlan. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36(3):318–324, 1987.

[40] J. Mitchell and S.G. Hall. Evaluating, comparing and combining density forecasts using the KLIC with an application to the bank of england and NIESR "fan" charts of inflation. *Oxford Bulletin of Economics and Statistics*, 67:995–1033, 2005.

[41] M. S. Paolella. Multivariate asset return prediction with mixture models. Swiss Finance Institute Research Paper No. 11-52, November 2011.

[42] B. Pesaran and M. H. Pesaran. Modelling volatilities and conditional correlations in futures markets with a multivariate t distribution. Cambridge Working Papers in Economics 0734, Faculty of Economics, University of Cambridge, 2007.

[43] S. Satchell. *Optimizing optimization : the next generation of optimization applications and theory.* Amsterdam [etc.] : Elsevier Ltd, 2010.

[44] E. Sentana and S. Wadhwani. Feedback traders and stock return autocorrelations: evidence from a century of daily data. *The Economic Journal*, 102(411):415–425, 1994.

[45] F. Sortino and L. Price. Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3):59–65, 1994.

[46] J. Tobin. Estimation of relationships for limited dependent variables. *Econometrica*, 26:24–36, 1958.

[47] Y. K. Tse and A.K.C. Tsui. A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2002.

[48] C.F.J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103, 1983.